# #GlockeAktiv: A corpus linguistic study of German youth language on YouTube

Louis Cotgrove

School of Cultures, Languages, and Area Studies

A thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy in Linguistics (Modern Languages)

April 2022

# Abstract

This thesis is a corpus linguistic investigation of the language used by young German speakers online, examining lexical, morphological, orthographic, and syntactic features and changes in language use over time. The study analyses the language in the *Nottinghamer Korpus deutscher YouTube-Sprache* ('Nottingham corpus of German *YouTube* language', or *NottDeuYTSch* corpus), one of the first large corpora of German-language comments taken from the video-sharing website *YouTube*, and built specifically for this project. The metadata-rich corpus comprises c.33 million tokens from more than 3 million comments posted underneath videos uploaded by mainstream German-language youth-orientated *YouTube* channels from 2008-2018.

The *NottDeuYTSch corpus* was created to enable corpus linguistic approaches to studying digital German youth language (*Jugendsprache*), having identified the need for more specialised web corpora (see Barbaresi 2019). The methodology for compiling the corpus is described in detail in the thesis to facilitate future construction of web corpora. The thesis is situated at the intersection of Computer-Mediated Communication (CMC) and youth language, which have been important areas of sociolinguistic scholarship since the 1980s, and explores what we can learn from a corpus-driven, longitudinal approach to (online) youth language. To do so, the thesis uses corpus linguistic methods to analyse three main areas:

1. Lexical trends and the morphology of polysemous lexical items. For this purpose, the analysis focuses on *geil*, one of the most iconic and productive words in youth language, and presents a longitudinal analysis, demonstrating that usage of *geil* has decreased, and identifies lexical items that have emerged as potential replacements. Additionally, *geil* is used to analyse innovative morphological productiveness, demonstrating how different senses of *geil* are used as a base lexeme or affixoid in compounding and derivation.
2. Syntactic developments. The novel grammaticalization of several subordinating conjunctions into both coordinating conjunctions and discourse markers is examined. The investigation is supported by statistical analyses that demonstrate an increase in the use of non-standard syntax over the timeframe of the corpus and compares the results with other corpora of written language.
3. Orthography and the metacommunicative features of digital writing. This

analysis identifies orthographic features and strategies in the corpus, e.g. the repetition of certain emoji, and develops a holistic framework to study metacommunicative functions, such as the communication of illocutionary force, information structure, or the expression of identities. The framework unifies previous research that had focused on individual features, integrating a wide range of metacommunicative strategies within a single, robust system of analysis.

By using qualitative and computational analytical frameworks within corpus linguistic methods, the thesis identifies emergent linguistic features in digital youth language in German and sheds further light on lexical and morphosyntactic changes and trends in the language of young people over the period 2008-2018. The study has also further developed and augmented existing analytical frameworks to widen the scope of their application to orthographic features associated with digital writing.

# Acknowledgements

Thank you, first and foremost, to my supervisors Professor Nicola McLelland and Dr Olivia Walsh, whose insight, feedback, and focus have been invaluable to all aspects of this thesis and have helped me develop as a researcher. I also extend my thanks to the Midlands4Cities Doctoral Training Partnership and the Arts and Humanities Research Council for the financial support to undertake this PhD.

Writing this PhD has been equal parts fun and stress. Thank you to my family who have provided unwavering encouragement at every stage of my thesis, food, and, above all, an internet connection from an early age. Thank you, danke, and gracias to all the various housemates who have put up with me over the past few years, I have really appreciated your support and the 'life' aspect of the work-life balance you have provided.

To the wonderful group of people who have passed through the doors of B57-B59 (and other offices) in the Trent Building: your friendships have been the highlight of these past four years (five with the Masters) – the academic discussions, not-so-academic discussions, advice, lunchtime gossip, parties, racquet sports, and camaraderie have made the PhD experience incredibly enjoyable. Thank you!

Lastly, my thanks go to the 1,633,311 different people who have left a comment under a German-language *YouTube* video between 2008 and 2018 that has found its way into the *NottDeuYTSch* corpus. Ohne euch wäre diese Doktorarbeit nicht möglich – danke.

# Notes on the thesis

## Linguistic Conventions

Linguistic conventions used in the thesis:

| Item | Convention | Example |
|---|---|---|
| Cited text | Quotation marks | "Example Quote" |
| Translated text | Single quotes | *Beispiel* ('Example') |
| Phonetic pronunciation | Square brackets | [ɪgˈzɑːmpl] |
| Graphemes | Corner brackets | < example > |
| Non-English words | Italics | *Beispiel* |
| Brand names | Italics | *YouTube* |
| Parts of speech | Square brackets with monospace typeface | `[NOUN]` |
| Coding examples | Square brackets with monospace typeface | `w[oi]l+[est]*n?` |

## Translations

Translations are provided underneath examples of language taken from the corpora used in this study where appropriate. All translations are by the author unless marked. Translations use British English terminology felt to be the closest in terms of retaining the flavour of the original, which often contains non-standard lexical items. Square brackets are used to provide brief descriptions for popular culture references following the original German, as well as standard spellings in extreme cases for readability, although non-standard orthography has mostly been left untouched and replicated in the English translations, where possible.

## Abbreviations

Abbreviations used in the thesis:

| Abbreviation | Explanation |
|---|---|
| DCK | *Dortmunder Chatkorpus*, The Dortmunder Chat Corpus |
| DECOW | The German Corpus of the Web 2016 B |
| DeReKo | *Das deutsche Referenz Korpus*, The German Reference Corpus |
| DiDi | The Digital Natives - Digital Immigrants Corpus of South Tyrolean CMC |
| DWDS-KK | *Digitales Wörterbuch der deutschen Sprache Kernkorpus des 20. Jahrhunderts*, The Digital Dictionary for the German Language Core Corpus of the 20th Century (1900-1999) |
| DWDS-KK21 | *Digitales Wörterbuch der deutschen Sprache Kernkorpus des 21. Jahrhunderts*, The Digital Dictionary for the German Language Core Corpus of the 21st Century (2000-2010) |
| DWDS-WK | *Digitales Wörterbuch der deutschen Sprache WebXL Metakorpus*, The Digital Dictionary for the German Language WebXL Metacorpus |
| IBK | *Internetbasierte Kommunikation und Social Media Korpora*, Internet-based Communication and Social Media Corpora |
| ipm | Instances per million words |
| ipmc | Instances per million comments |
| n = | Raw frequency in the *NottDeuYTSch* corpus |
| NottDeuYTSch | *Das Nottinghamer Korpus deutsche YouTube-Sprache*, The Nottingham Corpus of German YouTube Language |
| MoCoDa2 | The Mobile Communication Database 2 Corpus of WhatsApp Messages |
| SME | Social Media Entertainment (genre of *YouTube* content) |

## Citing from corpora

Examples taken from the *NottDeuYTSch* corpus are marked by only the year the comment was written as reference, e.g. Example 5.32 (2015). Examples taken from other corpora take the following format:

[Name of publication/Author(s)] ([Year]), [Corpus] [Reference code], e.g. Die Zeit (1969), DeReKo Z69/MAR.00187.

Examples taken from other scholarship take the following format:

[Author(s)] ([Year]: [Page]), e.g. Freywald (2016: 338).

# Contents

**1 Introduction** **1**

**Part I Investigating German youth language on YouTube** **7**

**2 Research context for Digitally Mediated Communication and youth language** **9**
2.1 Defining Digitally Mediated Communication . . . . . 11
2.1.1 Digital Writing . . . . . 14
2.1.2 Digitally Mediated Discourse . . . . . 28
2.2 The language of young people . . . . . 35
2.2.1 User-oriented approaches to youth language . . . . . 36
2.2.2 System-oriented approaches to youth language . . . . . 38
2.2.3 Lexical and morphological applications . . . . . 41
2.2.4 Syntactic applications . . . . . 43
2.2.5 Orthographic applications . . . . . 50
2.3 The language of young people in Digitally Mediated Communication . . . . . 51
2.4 The language of young people on *YouTube* . . . . . 53
2.5 Using *YouTube* comments to study digital youth language . . . . 54

**3 Constructing the *NottDeuYTSch* corpus** **59**
3.1 Selecting the data . . . . . 61
3.1.1 Principles of building the *NottDeuYTSch* corpus . . . . . 61
3.1.2 The identity of the commenters . . . . . 62
3.1.3 Identifying relevant *YouTube* channels . . . . . 64
3.2 Constructing the *NottDeuYTSch* corpus . . . . . 73
3.2.1 Extracting and cleaning the data . . . . . 73
3.2.2 Corpus representativeness and balance . . . . . 75
3.2.3 Determining the size of the *NottDeuYTSch* corpus . . . . 79
3.3 Statistical overview of the *NottDeuYTSch* corpus . . . . . 91
3.4 Applications of the *NottDeuYTSch* corpus . . . . . 95

**4 *Affengeil* and *abogeil* – lexical and morphological developments in digital youth language** **97**
4.1 *Geil* in history and scholarship . . . . . 99
4.1.1 Grammatical productivity of *geil* . . . . . 102
4.2 *Geil* in the *NottDeuYTSch* corpus . . . . . 108
4.3 Innovative compounding and derivation with *geil* . . . . . 109

4.3.1 Using the term 'affixoid' . . . 109
4.3.2 *Geil* used as an affixoid . . . 112
4.3.3 *Geil* as a base lexeme . . . 114
4.4 Use of *geil* over time in the *NottDeuYTSch* corpus . . . 134
4.4.1 Alternatives to *geil* in digital youth language . . . 138
4.5 The future of *geil* in digital youth language . . . 142

**Part II Syntactic developments in digital youth language: subordinating conjunctions 145**

**5 *Weil es kamen viele neue updates* – emergent non-standard word order following subordinating conjunctions 147**
5.1 Sampling the *NottDeuYTSch* corpus for syntactic analysis . . . 151
5.1.1 Selecting subordinating conjunctions . . . 154
5.1.2 Selecting 20 common verbs . . . 160
5.1.3 Capturing paratactic clauses in the *NottDeuYTSch* corpus 166
5.2 Analysing subordinating conjunctions and parataxis . . . 174
5.2.1 Parataxis in the *NottDeuYTSch* corpus . . . 175
5.2.2 Changes in parataxis over time . . . 183
5.2.3 Comparing parataxis across corpora . . . 185
5.3 Future trends for parataxis in digital youth language . . . 188

**6 From subordinating conjunctions to discourse markers? Investigating grammaticalization in digital youth language 191**
6.1 The role of the first field in paratactic clauses . . . 192
6.2 Grammaticalization of subordinating conjunctions to coordinating conjunctions and discourse markers . . . 199
6.2.1 Applying existing hypotheses of parataxis to the *NottDeuYTSch* corpus . . . 200
6.2.2 Distinguishing between coordinating conjunctions and discourse markers in the *NottDeuYTSch* corpus . . . 204
6.2.3 Modelling the grammaticalization of subordinating conjunctions . . . 211
6.3 Current and future trends for subordinating conjunctions . . . 217

**Part III Orthographic innovation in Digital Writing 221**

**7 *Deine rechtschreibung ist god like kappa :)* – the use of orthographic features for metacommunication 223**
7.1 Orthographic features in DMC . . . 224
7.1.1 Graphicons . . . 225
7.1.2 Creative orthography . . . 228
7.1.3 Metalinguistic signs . . . 230
7.2 Capturing orthographic features in the *NottDeuYTSch* corpus . . . 237
7.2.1 Identifying graphicons in the *NottDeuYTSch* corpus . . . 237
7.2.2 Identifying creative orthography in the *NottDeuYTSch* corpus . . . 239

7.2.3 Identifying metalinguistic signs in the *NottDeuYTSch* corpus 241
7.3 Analysing orthographic features in the *NottDeuYTSch* corpus . . 245
7.4 Developing a holistic framework of metacommunicative functions in Digital Writing . . . . . 246
7.4.1 Illocutionary force . . . . . 248
7.4.2 Indexing identity . . . . . 253
7.4.3 Illustration . . . . . 257
7.4.4 Structuring the information within the digital space . . . . 263
7.5 Future applications of the metacommunication framework . . . . 270

**8 Conclusions 273**

**Part IV Bibliography and Appendix 281**

**Bibliography 283**

**A Appendix: Statistical tables of the *NottDeuYTSch* corpus 343**

# List of Tables


2.1 Examples of the combinations between conception and medium based on the Proximity-Distance model. Adapted from Kilian (2010: 69) . . . 17
2.2 Updated parameters of the Proximity-Distance model. Adapted from Koch and Österreicher (2007: 351) . . . 18

3.1 Breakdown of sources used to identify channels to include in the *NottDeuYTSch* Corpus . . . 72
3.2 List of channels included in the *NottDeuYTSch* corpus . . . 74
3.3 Number of videos and associated statistics for each *YouTube* category from the *YouTube* channels in the database . . . 77
3.4 Total number of videos per category and year in the pre-corpus database . . . 78
3.5 Number and percentage of comments under videos in each category and year in the pre-corpus database . . . 81
3.6 Projected number and percentage of comments under videos in each category and year in the *NottDeuYTSch* corpus . . . 82
3.7 Number of videos selected for the *NottDeuYTSch* corpus per category and year . . . 85
3.8 'Gaming, 2012' Intersection showing totals and proportions of videos and comments uploaded and posted to each channel . . . 87
3.9 'Gaming, 2012' Intersection showing how many videos and comments are to be selected from each channel . . . 87
3.10 Eight selected videos from the PietSmiet channel in the 'Gaming, 2012' intersection, showing the F values, target number of comments, and actual number of extracted comments . . . 89
3.11 Videos selected for 'Gaming, 2012' intersection with number of comments . . . 90
3.12 Statistical overview of the *NottDeuYTSch* corpus . . . 91
3.13 Number of comments per video category and year in the *NottDeuYTSch* corpus . . . 93
3.14 Projected number and percentage of comments under videos in each category and year in the *NottDeuYTSch* corpus . . . 94

4.1 Grammatical categories of *geil*, adapted from Mroczynski (2018: 331) . . . 108
4.2 Frequencies of the three attested senses of *geil* in the *NottDeuYTSch* corpus . . . 108
4.3 The ten most frequent base lexemes with *geil* as a suffixoid in the *NottDeuYTSch* corpus . . . 114

4.4 Frequencies of German youth language prefixoids in Androutsopoulos (1998: 106-108) combined with *geil* in the *NottDeuYTSch* corpus . . . 118
4.5 The ten most frequent intensifying prefixes and prefixoids with *geil* in the *NottDeuYTSch* corpus . . . 120
4.6 Ten most frequent morphological intensifiers in *geil* constructions containing multiple intensifiers in the *NottDeuYTSch* corpus 122
4.7 Most frequent pairs of consecutive prefixoids in *geil* constructions in the *NottDeuYTSch* corpus . . . 128
4.8 Most frequent triplets of consecutive prefixoids in *geil* constructions in the *NottDeuYTSch* corpus . . . 128
4.9 A selection of constructions with a large number of prefixoids preceding *geil* in the *NottDeuYTSch* corpus . . . 130
4.10 Most frequent suffixes with *geil* in the *NottDeuYTSch* corpus . . . 131
4.11 Frequency of positive expressions in the *NottDeuYTSch* corpus listed in alphabetical order (instances per million) . . . 138
4.12 Frequency of positive expressions in 2009 (left), 2013 (centre), and 2018 (right) in the *NottDeuYTSch* corpus . . . 140

5.1 The topological model of German syntax, adapted from Leibniz-Institut für Deutsche Sprache (2017) . . . 148
5.2 Topological field diagram for a hypotactic clause . . . 149
5.3 Topological field diagram for a paratactic clause . . . 149
5.4 Topological field diagram for a 'subordinating' conjunction with paratactic word order . . . 149
5.5 Explanation of an example Regular Expression . . . 153
5.6 Example comments from the *NottDeuYTSch* corpus of paratactic word order following the selected conjunctions . . . 160
5.7 The 20 most frequent verbs (finite forms) from the selected *DWDS* corpora and the 20 selected verbs for analysis of the *NottDeuYTSch* corpus . . . 163
5.8 Frequency of the selected conjunctions that introduce paratactic clauses in the *NottDeuYTSch* corpus . . . 176
5.9 Example comments from the *NottDeuYTSch* corpus of paratactic word order following the selected conjunctions . . . 179
5.10 Potential multilingual emergent grammar structures in the *NottDeuYTSch* corpus . . . 182
5.11 Comparison of the rates of parataxis across selected corpora . . . 186

6.1 First field constituents with the largest disparity between their frequencies in paratactic *weil* clauses and all *weil* clauses in the *NottDeuYTSch* corpus . . . 195

7.1 Types of graphicon . . . 226
7.2 Complex inflectives, adapted from Schlobinski (2001: 193) . . . 232
7.3 Orthographic features in Digital Writing that fulfil metacommunicative functions . . . 244

A.1 List of channels included in the *NottDeuYTSch* corpus . . . 344

A.2 Number of videos selected for the *NottDeuYTSch* corpus per category and year . . . . . 345
A.3 Number of comments per video category and year in the *NottDeuYTSch* corpus . . . . . 346
A.4 Statistical overview of the *NottDeuYTSch* corpus . . . . . 347

# List of Figures


2.1 The six new punctuation marks developed by Herve Bazin (1966: 142) . . . 16
2.2 Proximity-Distance model (Koch and Oesterreicher 1985: 23) . . 17
2.3 Enriched communicative model (Landert & Jucker 2011: 1427) . 19
2.4 Expansion of Koch and Oesterreicher's Proximity-Distance Model incorporating synchronicity of medium (Dürscheid 2003: 13) . . 20
2.5 Examples of Bitmoji, from Snapchat Inc. (2021) . . . 26
2.6 Three states of the 'Unicorn' animoji . . . 26
2.7 Model of linguistic innovation (Kotsinas 1997), taken from Androutsopoulos (2005: 182) . . . 39
2.8 Model of linguistic innovation, adapted from Bahlo (2021a) . . . 40

3.1 Left: BRAVO magazine cover, 29.08.2014, with Bianca Heinicke (Bibi) (top centre), Y-titty (bottom right), and ApeCrime (bottom left). Right: 'BRAVO Tubestars' cover, 29.10.2014, featuring popular *YouTubers* . . . 67
3.2 1LIVE Babo-Bus: *YouTube* video description contains links to other online platforms belonging to the uploader . . . 69
3.3 *YouTube* channel 'About' page, showing channel description, number of subscribers, the date that the channel was created, related channels (curated by the channel owner), and the total number of views of all the videos in the channel . . . 73

4.1 Frequency of *geil* in the *DWDS Kernkorpora* (These data do not include online sources of language) . . . 102
4.2 Frequency of *geil* in three newspapers (*Die Tageszeitung* - TAZ, *Mannheimer Morgen* - MAMO, and *Die Zeit* - ZEIT) from 1986-2003, taken from Androutsopoulos (2005: 192). Frequencies in ipm. . . . 103
4.3 Changes in frequency of *hammergeil* over time in the *NottDeuYTSch* corpus . . . 120
4.4 Distribution of morphological intensifiers in *geil* constructions . . 123
4.5 Distribution of number of prefixoids with *affen* in *geil* constructions in the *NottDeuYTSch* corpus . . . 127
4.6 Changes in frequency of *geil* over time in the *NottDeuYTSch* corpus 135
4.7 Changes in frequency of different meanings of *geil* over time in the *NottDeuYTSch* corpus . . . 137
4.8 Changes in frequency of positive expressions in the *NottDeuYTSch* corpus . . . 141

5.1 Frequency of part-of-speech classification of the excluded uncommon conjunctions in the *DeReKo* (red) and *DWDS Webkorpus* (blue) . . . 155
5.2 A comparison of personal pronoun frequency with mental content verbs in the *DWDS Kernkorpus 21*, *Dortmunder Chat-Korpus* and *NottDeuYTSch* corpus . . . 167
5.3 Frequency of paratactic clauses following the selected conjunctions in the *NottDeuYTSch* corpus over time . . . 183
5.4 Continuum of syntactic characteristics based on language formality continuum with a visual indication of the formality of the selected corpora . . . 187

6.1 Linear model of grammaticalization of subordinating conjunctions, adapted from Gohl and Günthner (1999: 70) . . . 212
6.2 Refined model of grammaticalization of subordinating conjunctions . . . 213

7.1 Usage of emoji and emoticons over the timespan of the *NottDeuYTSch* corpus . . . 228
7.2 Mocking Spongebob Meme with accompanying stylised text . . . 231
7.3 Frequency of inflectives over time in the *NottDeuYTSch* corpus . . . 232
7.4 The 'gronkhHi' *Twitch* emote . . . 236
7.5 The 'kappa' *Twitch* emote and variants: kappaPride and kappaClaus 236
7.6 ASCII art of a person posing or dancing, 2016 . . . 241
7.7 BonnyTrash homepage banner . . . 257
7.8 Ludic illustration using emoji, from Dürscheid and Siever (2017: 270) . . . 264
7.9 ASCII art of a person dancing, 2016 . . . 266
7.10 ASCII art of a tank and text, 2014 . . . 266
7.11 *YouTube* comment expanded and with "read more" button . . . 268
7.12 The four metacommunicative functions expressed using a single orthographic feature . . . 271

# 1

# Introduction

This thesis is a corpus linguistic investigation of lexical, morphological, syntactic, and orthographic features of the language written by young German speakers in comments below videos on *YouTube*, the video sharing website and second-most visited site in the world after *google.com* (as of 2021, Alexa Internet, Inc. 2021). The study is situated at the intersection of Digitally Mediated Communication (DMC) (also referred to as Computer-Mediated Communication) and German youth language,[1] two areas of sociolinguistic scholarship, which began in the 1980s. Research at the intersection of these fields, which I refer to as digital youth language, has primarily focused on forums and social media sites, (e.g. *bravo.de*, Dürscheid 2006; and *Facebook*, Androutsopoulos 2015), or messaging services, (e.g. *ICQ*, Kleinberger Günther & Spiegel 2006; and *WhatsApp*, Dürscheid & Frick 2014).

However, the linguistic features used by young people in *YouTube* comments have rarely been analysed in studies of either DMC or youth language, despite *YouTube* becoming one of the most-used online sites of communication in this

[1] Also referred to as *Jugendsprache* – see Chapter 2, Section 2.1 for a discussion of these terms.

demographic (Saferinternet.at 2018), with 86% of 12-19-year-olds reporting that they regular watched *YouTube* videos in 2018 (Bahlo et al. 2019: 80). To address this underdeveloped field of scholarship, I have constructed the *Nottinghamer Korpus deutscher YouTube-Sprache* ('Nottingham corpus of German *YouTube* language', or *NottDeuYTSch* corpus) specifically for this study to investigate the language of young German-speakers in digital spaces.

The *NottDeuYTSch* corpus is a collection of over 33 million words written between 2008 and 2018 taken from the comment sections of 112 mainstream German-language *YouTube* channels that produce content targeted at young people (for an in-depth explanation of the corpus, see Chapter 3). While other corpora of digital German language have been constructed, they have focused on other sources of data, e.g. websites and online forums (the *DECOW* corpus, Schäfer 2015; *DWDS WebXL Korpus*, Geyken et al. 2017; Barbaresi & Geyken 2020), South Tyrolean *Facebook* texts (the *DiDi Korpus*, Glaznieks & Frey 2020), Internet Relay Chat (IRC) messages from students (the *Dortmunder Chat Korpus*, Beißwenger et al. 2015), *WhatsApp* messages (the *MoCoDa2* corpus, Beißwenger et al. 2020), and SMS, e-mail, IRC, Twitter, and Wikipedia article and discussion pages (the *IBK und Social Media-Korpora*, Lüngen & Kupietz 2020).

Some of the corpora aim to capture a wide range of DMC text types, but the majority are highly specialised: either focusing on one method of communication or on one target group. The range of specialised corpora demonstrates the "unparalleled and rapidly evolving diversity in terms of speakers and settings" in DMC (Barbaresi 2019: 29), although none of the above-mentioned corpora have exclusively focused on the language of young people. Indeed, Barbaresi (2019: 30) advocates for the creation of more specialised corpora of online language, "to complement existing collections, as they allow for better coverage of specific written text types and genres, especially the language evolution seen through the lens of user-generated content, which gives access to a number of variants, socio- and idiolects". Androutsopoulos & Tereick (2016: 366–67) also advocate specifically for more linguistic research using *YouTube*, highlighting "comment interaction, remix and multimodality, discourse participation, perfor-

mance and stylization of linguistic variability" as potential areas of study.

The *NottDeuYTSch* corpus answers these calls, providing an unparalleled opportunity for exploratory study of the colloquial Digitally Mediated Communication of and between young people. The period covered by the corpus, 2008-2018, sits within the internet epoch referred to as Web 2.0 (O'Reilly & O'Reilly 2005), an era of online and digital communication that began in the mid-2000s characterised by "social interaction and user-generated content", rather than monological, information repositories (Herring 2013: 1). This decade was also an important period of technological transition from PC to mobile-based communication for many young people, who experienced the "digitalisation [of their] everyday lives" (Döring 2010: 161), acquiring personal access to the internet through smartphones, rather than being restricted to family or school computers or internet cafes. The corpus therefore can potentially capture any linguistic changes in digital youth language that may have accompanied the technological changes.

The language of young people online and possible changes over time comprise the central research question of this study: what can we learn from a corpus-driven, longitudinal approach to digital youth language? Supporting the central question are three further research questions, which each focus on a different linguistic aspect in more detail:

I. How can corpus linguistic methods be used to investigate popular lexical items and lexical trends and developments? As part of those lexical developments, how can corpus linguistic methods be used to investigate morphological differences in polysemous lexical items?
II. To what extent have non-standard syntactic phenomena developed over the time span of the *NottDeuYTSch* corpus and how can corpus linguistic methods be used to analyse them?
III. How are orthographic strategies used to alter the metacommunicative functions of digital messages and how can the *NottDeuYTSch* corpus be used to provide data to develop a framework for future analysis of Digital Writing?

The thesis is split into three parts, I to III, with each addressing the corresponding research question listed above. Part I of the thesis, containing Chapters 2 to 4, provides the academic context in which all the research questions are studied and introduces the potential of the corpus to analyse digital youth language through a case study of lexis. Chapter 2 introduces the concepts of Digitally Mediated Communication and youth language. The chapter presents an overview of linguistic scholarship in the field of Digitally Mediated Communication, particularly written forms, referred to as Digital Writing. The chapter then examines previous linguistic and ethnographic research on German-speaking young people, charting historical developments in attitudes towards youth language, and the interaction between youth language and standard language. Finally, the chapter discusses the little-researched area of digital youth language, which sits at the intersection of DMC and German youth language research, specifically focusing on historical and technological developments, culminating in an analysis of the few existing studies on German youth language within *YouTube*.

Chapter 3 explains the construction of the *NottDeuYTSch* corpus, defining the guiding principles and identifying possible ethical concerns of the corpus. The chapter clarifies the selection process of the data sources, justifying the parameters for the selection of *YouTube* channels, videos, and comments, as well as determining the optimal size of the corpus to enable satisfactory investigation of the research questions. The chapter also outlines the processes to extract and collate the comments, as well as cleaning the data. Finally, a statistical overview of the *NottDeuYTSch* corpus is provided.

Chapter 4 then analyses the lexical and morphological development of the polysemous and highly productive adjective, *geil*, presenting a longitudinal study of the different senses in which it is used in the *NottDeuYTSch* corpus. The analysis demonstrates that, contrary to its status as one of the most iconic words in German youth language, usage has dramatically decreased between 2008 and 2018 and the corpus is used to identify lexical items that have emerged as potential replacements. The chapter then specifically analyses the innovative morphological productiveness of *geil*, demonstrating how different senses of the word are used either as a base lexeme or as an affixoid in compounding

and derivation.

Part II, comprised of Chapters 5 and 6, investigates syntactic developments in digital youth language, through a case study that analyses seemingly non-standard subordinating clauses, demonstrating novel grammaticalization of subordinating conjunctions, which allow them to function as both coordinating conjunctions and discourse markers. The chapter also shows that these phenomena are increasing in frequency over the time frame of the corpus.

In Part III, Chapter 7 develops a framework to examine metacommunicative functions in digital youth language, i.e. how extra information is encoded in a message outside of lexical choice, expressed through orthographic variation and features, such as the repetition of letters, spacing, or the use of certain emoji. The framework unifies previous scholarship on individual features to integrate a wide range of metacommunicative strategies within a single system of analysis.

Bahlo et al. (2019: 96) stated that

> tracing processes of linguistic variation and change against the background of (globally) networked online communities is certainly one of the great challenges for (diachronic) sociolinguistics in the near future. The desideratum for youth language research is to clarify the role of users in the phase of adolescence within these processes (e.g. as creative drivers of innovation or as multipliers).

Through the application of qualitative and quantitative corpus linguistic methodologies to the *NottDeuYTSch* corpus, this study engages with that challenge: it uncovers emergent lexical, morphological, and syntactic features and developments in digital youth language over the period of 2008-2018, develops existing analytical frameworks to widen the scope of their application to orthographic features associated with digital writing, and, above all, highlights the innovation, creativity, and diversity in the language of young people.

# Part I

# Investigating German youth language on YouTube

# 2

# Research context for Digitally Mediated Communication and youth language

This chapter presents an analysis of the key scholarship and theoretical frameworks in the areas of both Digitally Mediated Communication (DMC) and youth language, as well as the significantly smaller research area at the intersection of these two fields. This analysis of previous studies contextualises the corpus linguistic investigations of lexical, morphological, syntactic, and orthographic features of digital youth language set out in this thesis. As Hugger (2014: 11) has stated, "Jugend ist gegenwärtig nicht nur Offline-Jugend, sondern zugleich Online-Jugend" ('youth today is not just offline youth, it is simultaneously online youth'), and it is therefore important to acknowledge how the fields of DMC and youth language are intertwined when analysing the linguistic phenomena that are the focus of this thesis.

Section 2.1 examines the scholarship on communication in digital environments, e.g. the use of mobile phones, computers, and other internet-enabled devices to interact. The section analyses the competing terminology used to refer to

communication using digital devices and presents an argument for the use of the term 'Digitally Mediated Communication' to refer to such communication, which is used within this thesis. The section then investigates the history and trends within the field since the 1980s, concentrating on two areas of research in a German context: the linguistic features of DMC, and the interactive and discursive practices characteristic of DMC. Scholarship on the linguistic features of DMC contextualises the investigation of non-standard syntax in Part II, comprising Chapters 5 and 6, in particular, the comparative analyses of the *NottDeuYTSch* corpus with other corpora of written language. Scholarship on both the linguistic and interactive aspects of DMC provides the context for the analyses of lexical and morphological features and longitudinal changes in the analysis chapter of Part I of this thesis (Chapter 4), as well as the analysis of digital orthographic strategies for metacommunication in Chapter 7.

Section 2.2 reviews research on language used by young people in German-speaking contexts, particularly the use of youth-specific sociolects, referred to collectively as *Jugendsprache*. The section organises and analyses the scholarship according to the two main methodological approaches, which inform different aspects of my analyses of the *NottDeuYTSch* corpus. I refer to the first approach as 'user-oriented analysis' and it focuses on the relationship between youth language and standard language in German-speaking contexts and the construction of identities through language. This focus on identities and language choice informs my analyses of metacommunicative functions in Chapter 7, as well as lexical developments in Chapter 4. I refer to the second approach as 'system-oriented analysis' and this focuses on the influences and systemic processes behind the adoption of certain linguistic features in youth language. This approach particularly informs the analysis of the syntactic developments in digital youth language covered in Chapter 5, and it also contributes to the longitudinal lexical analysis presented in Chapter 4.

Section 2.3 reviews work specifically on the language used by young people in Digitally Mediated Communication, highlighting the overlaps and differences between the two research areas in this relatively new field, particularly work on the German language. Section 2.4 then narrows the focus to research that

has analysed the language in *YouTube* videos and comments, particularly the language used by young people, identifying the current gaps in the body of knowledge and methodologies that this study addresses.

## 2.1 Defining Digitally Mediated Communication

The most widely-used term to refer to human-to-human communication through digital media has been 'Computer-Mediated Communication' (CMC), a term popular since the 1980s, which replaced the earlier term, "computerized conferencing", to describe "any system that uses the computer to mediate communication among human beings" (see Hiltz & Turoff 1993 [p. xix], originally published in 1978). The initialism CMC gained traction as new, informal methods of communication became available, such as online message boards. However, since the mid-2000s, it has been argued that CMC is no longer appropriate for at least three reasons:

1. Communication using microprocessor-based technology is no longer purely keyboard-based; it encompasses audial, visual, and audio-visual communication, and is often multimodal, using combinations of text, image, and audio (see Jucker & Dürscheid 2012: 4-8), as well as even haptic feedback (i.e. vibrations).
2. CMC practices can be performed on devices that are no longer considered 'computers', such as mobile phones, tablets, or wearable technology (see Carr 2020).
3. Linguistic features associated with CMC occur in communication using non-computer, and non-internet devices, such as sending an SMS using a mobile phone (see Herring 2007).

Herring (2007) partially addressed points 2. and 3. by expanding the definition of CMC to "text-based human-human interaction mediated by networked computers or mobile telephony", but the focus on text-based communication excludes the other modes mentioned in point 1., such as audio-based technology. Other suggestions have included "electronic language" (Collot & Belmore 1996: 13), "electronically-mediated communication" (Baron 2008: xii),

"internet-mediated communication" (Yus 2011), and "electronic communication" (Herring 2012), and even simply "Mediated Communication" (Carr 2020). Carr (2020: 10) suggests that there is a wider epistemological problem with the term CMC, as "so much of the communicative experience is mediated, often by omnipresent digital tools" and any term must be inclusive of the wide range of devices used for communication, as opposed to "technology anchored to a desk by a cord" (Carr 2020: 10). The use of the term "Mediated Communication" (MC), therefore, de-emphasises the role of computers and emphasises the "process of mediation". For future scholarship, Carr (2020: 16) calls for a holistic analytical approach, to "theorize MC analogically [...] rather than trying to idiosyncratically consider each new technological development". While Carr's refocusing of scholarship is useful for developing future methodological approaches, it is reliant on a "technology-agnostic approach" (Carr 2020: 17), which assumes that the language can be separated from the medium. However, human-to-human interaction can vary even between similar platforms (e.g. the mid-2000s instant messaging clients, MSN messenger and ICQ), let alone sites with differing architecture, which casts doubt on the usefulness of this approach. While linguistic features can sometimes be generalised across all communicative technologies, the platform is often a factor that has to be taken into account, regardless of whether it is a main focus within the study. The term 'Mediated Communication' is arguably more future-proof than simply replacing the 'computer' in 'computer-mediated communication' with an alternative technological term common to the devices, such as 'microprocessor', as technological advancements could also render this obsolete. However, Mediated Communication is very broadly interpretable, as air, water, and copper coil are also media through which we communicate, making the term open to very broad interpretation, which reduces its usefulness. In this study, therefore, I adopt the term 'Digitally Mediated Communication' (henceforth referred to as DMC) (Yao & Ling 2020; analagous to the German, 'digital vermittelte Kommunkation' Androutsopoulos 2020a: 137), as it is device-agnostic, whilst still making the technologically-mediated nature of the communication explicit, unlike other terminological suggestions that contain 'digital', e.g. "digital com-

munication" (Tagg 2015; Georgakopoulou & Spilioti 2016).

In this thesis, I often refer to specific communicative aspects of DMC, each requiring their own separate terminology: the analysis of the features of typed communication (i.e. 'graphically encoded,' Jucker & Dürscheid 2012: 3) is referred to as 'Digital Writing'; and separately, the analysis of discourse-oriented processes in DMC, referred to as 'Digitally Mediated Discourse'. The two concepts are explained more fully below, and the scholarship falling under the two aspects is explored in two separate sections, 2.1.1 and 2.1.2:

**Digital writing (Section 2.1.1):** To analyse scholarship on the lexical, orthographic, and morphosyntactic features pertaining to written (typed) DMC (as opposed to spoken language), the term 'Digital Writing' is preferred, following Androutsopoulos (2018: 741–42). Other earlier suggested terms have included "schriftbasierte Chat-Kommunikation" (written chat communication) (Beißwenger & Storrer 2012) and "keyboard-to-screen communication" (Jucker & Dürscheid 2012), but the scope of these terms is restricted, either to one form of communication, i.e. chat, or to the use of a keyboard for input, which does not account for other graphical input methods, such as pen tablets and keypads. For the purpose of this study, I define 'Digital Writing' as the transmission of graphical characters and icons (e.g. emoji) within human-to-human interaction (following Herring 2007: 13-14). Such a definition also includes text and characters generated using speech-to-text tools or other accessibility features, but it excludes computer-generated language, human-computer interaction, or computational linguistic methodologies that simulate Digital Writing, e.g. Natural Language Processing (NLP), Markov chains, Turing tests, chat bots, and other processes using Artificial Intelligence.

**Digitally Mediated Discourse (Section 2.1.2):** Previous scholarship on discourse-oriented DMC has used the terms "Computer-Mediated Discourse" (CMD) (Herring 1992, 2007), "interactive written discourse" (Ferrara, Brunner, & Whittemore 1991), "digital discourse" (Thurlow & Mroczek 2011), and "digital networked writing" (Androutsopoulos 2011: 145). Following the reasons

discussed for Computer-Mediated Communication above, the term 'Computer' is not appropriate, so as a compromise between the above terms, 'Digitally Mediated Discourse' (DMD) is used. The term encompasses a wide range of aspects of DMC, which are centred around language use and interaction between users. Digitally Mediated Discourse scholarship includes research that has focused on dialogue, e.g. conversation structure and turn-taking, identity work, e.g. the construction of social identities (such as gender and ethnicity), including language choice, translanguaging, multilingualism, and power dynamics, as well as the construction of digital communities, e.g. communities of practice, speech communities, and other support networks (see Herring 2019: 39-40).

Having defined the terminology that is used to analyse the linguistic features in the *NottDeuYTSch* corpus, I turn now to an examination of previous scholarship on Digital Writing and Digitally Mediated Discourse.

### 2.1.1 Digital Writing

Novel linguistic practices have been a central focus of Digital Writing scholarship from the very beginning of the field in the late 1970s. For example, Carey (1980), identified emergent orthographic methods of communicating emotion when using "computer conferencing systems", such as "vocal spelling" (the repetition of graphemes to represent prosody). Many formal registers of Digital Writing, such as those found in online newspapers and reference sites, closely resemble their offline counterparts in using standard written language. However, increased access to digital communication methods in the late 1990s and early 2000s saw a huge increase in the Digital Writing of ordinary users, and this area of language was often the focus in studies at this time. These examples of Digital Writing often contained language forms considered to represent or emulate spoken language, and were referred to as "typed conversations" (Storrer 2001), "typed dialogue" (Dürscheid & Brommer 2016) or "written colloquial speech" (Kilian 2001). Most of the research, in English and German, highlighted the "stylistic distinctiveness" (Crystal 2006: 31) and "colloquial" lin-

guistic features of Digital Writing (Hentschel 1998: 3.2.1), such as abbreviations and acronyms (e.g. Werry 1996: 53-56; Kilian 2001), obscenities (e.g. Diekmannshenke 2000: 144), non-standard syntax (Kilian 2001), and emoticons, such as < :) > (e.g. Runkehl, Schlobinski, & Siever 1998: 11-12; Wolf 2000). In studies of German in particular, a significant proportion of scholarship was devoted to the use of "inflectives", verb stems that are enclosed in punctuation to represent actions, e.g. < *grins* >, which indicates that the author of the comment would be grinning, were the conversation to take place face-to-face (see Hentschel 1998; Runkehl, Schlobinski, & Siever 1998; Schlobinski 2001, 2005; Kilian 2001; Androutsopoulos & Schmidt 2002; Bader 2002; Wirth 2005; Siever 2006; Dürscheid, Wagner, & Brommer 2010; Herring 2012; Storrer 2013).

However, the use of orthographic features in written language is both more complex and older than the digital era. Vellusig (2000) provided evidence of such features in private correspondence in the 18th century, described as "a written form of speech [...], which emulates the gestures of oral interaction through writing" (Vellusig 2000: 30). These orthographic features included the use of phonetic spellings, interjections, direct questions, the creation of neologisms, and the emulation of spoken Latin rhetorical style. More recently, in the 1960s, there were several attempts to popularise characters that could convey suprasegmentals of spoken language: stress, tone, pace, and volume (Beeman 2019: 2). For example, the 'interrobang', either written as < ⁉ > or < ?! >, was developed by Martin Speckter to convey "surprised rhetorical questions", and six new punctuation marks were suggested by Bazin (1966: 142) to convey certain tones, such as irony and doubt (see Figure 2.1), as part of his wider attempts at orthographic reform of the French language.

Following an initial wave of DMC research that tended to focus on discursive strategies, most notably gendered power dynamics (e.g. Curtis 1992; Kantrowitz & Rosenberg 1994; Tannen 1994; Herring 1996c, 1996b), orthographic features were often the focus of studies from the mid-1990s, covering both synchronous (e.g. Internet Relay Chat, SMS, and messaging clients) and asynchronous communication (e.g. newsgroups, forums, and email). However, the rise of digital communication and the subsequent increase in the digital

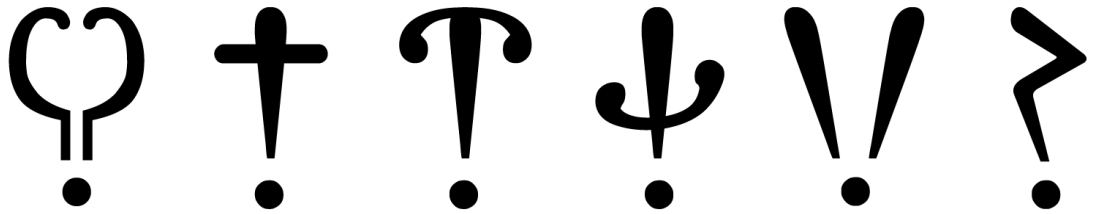



Figure 2.1: The six new punctuation marks developed by Herve Bazin (1966: 142)

written word, particularly in digital youth language, has seen considerable innovation and creativity in the development of orthographic features. This means that attempts to collate such features, e.g. Carey (1980) and Crystal (2004), may quickly become outdated. While some features, such as shouting capitals, have endured, other styles or customs in digital writing fall out of use and new ones replace them. Chapter 7 addresses this short 'half-life' of orthographic features in Digital Writing, suggesting how to categorise them so that analysis can remain 'future-proof'.

#### 2.1.1.1 Frameworks to analyse Digital Writing

The general view of early scholarship that DMC was "neither simply speech-like nor simply written-like" (Yates 1996: 46) was developed within German-language research, which analysed DMC using the framework of "orality" and "literality" (*Mündlichkeit* and *Schriftlichkeit*), i.e. the relationship between spoken and written language (see Söll & Hausmann 1980; Ong 1982; Koch & Österreicher 1985). Koch & Österreicher (1985, later refined in 2007) developed the influential *Nähe-Distanz Modell* ('Proximity-Distance model'), shown in Figure 2.2, which has been used since the 1990s to examine Digital Writing (e.g. Günther & Wyss 1996; Schlobinski 2005; Beißwenger & Pappert 2020). The model posits that spoken (oral) and written (graphic) language sits at either end of a spectrum, each associated with a set of characteristics that comprise their 'conception'. That is, conceptually 'oral' language is dialogical, expressive, and spontaneous, and conceptually 'literal' language is monological, objective, and reflected. However, the 'conception' is separate from the medium, whether the

language is produced phonically or graphically. Although conceptually oral features are typically associated with phonic language production, they could also be present in written language, and similarly, conceptually literal features could be present in phonic language. Examples of each of the four combinations between concept and medium are presented in Table 2.1. Due to the variety of conceptually oral features in Digital Writing it was widely considered that DMC sat somewhere between "literality and orality" (Bader 2002).

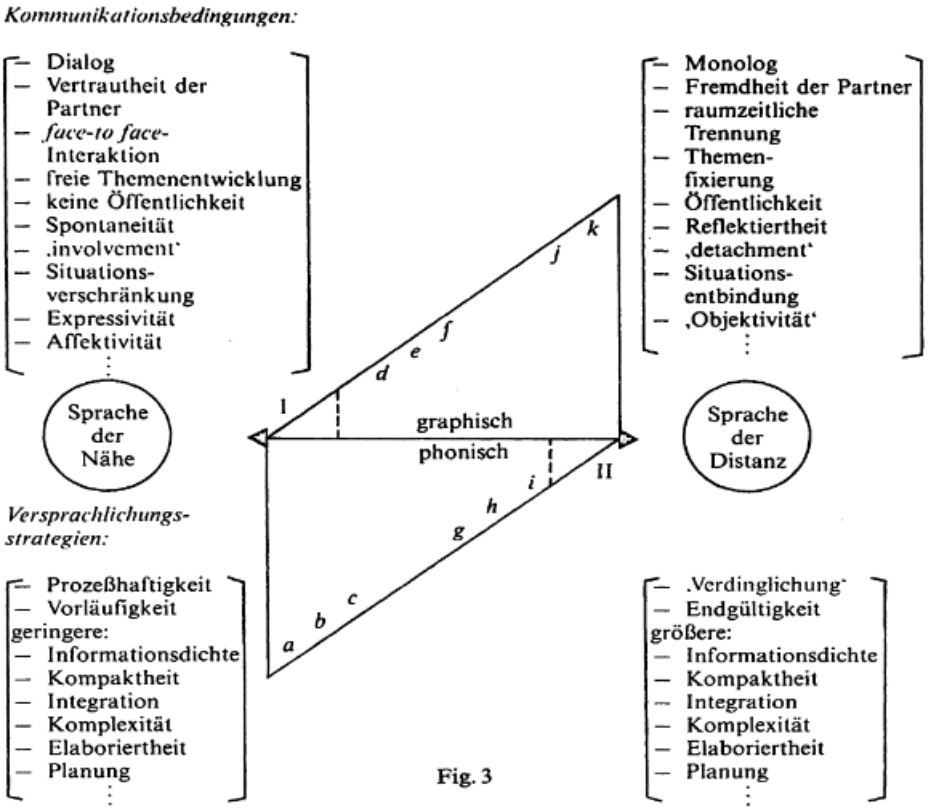



Figure 2.2: Proximity-Distance model (Koch and Oesterreicher 1985: 23)

Table 2.1: Examples of the combinations between conception and medium based on the Proximity-Distance model. Adapted from Kilian (2010: 69)

| | | **Medium** | |
|---|---|---|---|
| | | Graphic | Phonic |
| **Conception** | Oral | <nehm wa mal an> | [ne:mwɐma'an] |
| | Literal | <nehmen wir einmal an> | [ne:mən‿wɪɐ‿aɪmal‿an] |

In the years since the original publication of the *Nähe-Distanz Modell*, there has been significant discussion regarding the application of the model to DMC, and several revisions have been proposed. Ágel & Hennig (2006) criticised the

division of the two conceptions into "communicative conditions" and "verbalisation strategies" in the original model (as illustrated in Figure 2.2), calling them unviable to facilitate "a reliable classification of individual types of discourse in the *Nähe-Distanz* continuum" (Ágel & Hennig 2006: 183). To address this, Ágel & Hennig (2006) developed their own model that reframed and expanded the conditions and strategies of Koch & Österreicher (1985), dividing them into 25 intersecting levels and parameters, with around 140 separate features that distinguished *Nähe* ('proximity') and *Distanz* ('distance'). The Ágel & Hennig (2006) model examined gave an impressive level of granularity, and was, in part, later adopted by Dürscheid, Wagner, & Brommer (2010) in a study of online youth language. Responding to Ágel & Hennig (2006), Koch & Österreicher (2007: 351) combined their original conditions and strategies into a single group of ten pairs of "communicative parameters", which are shown in Table 2.2.

Table 2.2: Updated parameters of the Proximity-Distance model. Adapted from Koch and Österreicher (2007: 351)

| **Proximity** | **Distance** |
|---|---|
| Private | Public |
| Familiarity with conversational partner | Unfamiliarity with conversational partner |
| strong emotional involvement | low emotional involvement |
| Influenced by situation and/or actions | Disassociation from situation and/or actions |
| Referential proximity | Referential distance |
| Spatio-temporal proximity (face-to-face) | Spatio-temporal distance |
| Communicative cooperation | No communicative cooperation |
| Dialogicity | Monologicity |
| Spontaneity | Reflectedness |
| Unrestricted evolution of topic/theme | Fixed topic/theme |

However, Landert & Jucker (2011) suggested that the *Nähe-Distanz Modell* needed further revisions to account for language proximity differences based on the privacy or publicness of the topic and context of the communicative act. In their revised model, based on their comparison of letters to newspaper editors and comments under online newspaper articles (see Figure 2.3), a private topic (e.g. talking about breakfast) in a private context (e.g. a chat at

home) is more likely to contain the "language of immediacy", compared to a "non-private" topic in a public context (e.g. a scientific article) (Landert & Jucker 2011: 1427). They identify considerable differences in language between online comments and letters to the editor, arguing that online comments are more "conceptually oral" because they discuss private topics. As *YouTube* comments sections are also publicly accessible areas, we can expect that they will also exhibit the language of immediacy. However, Landert & Jucker (2011) do not take into account the accessibility of publishing the communication, which only receives a brief mention within the article. There are relatively few barriers to posting a comment, whereas letters to the editor are selected and edited before publication.

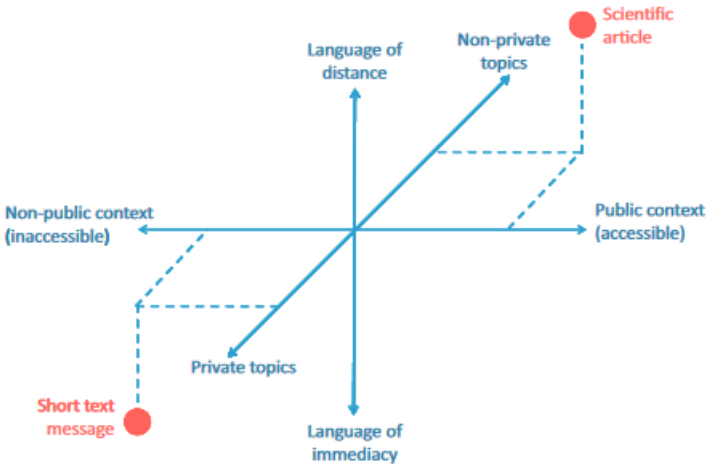



Figure 2.3: Enriched communicative model (Landert & Jucker 2011: 1427)

To accommodate the many varieties of DMC, Dürscheid (2003) and Schlobinski (2006a) proposed sub-dividing Koch and Österreicher's continuum between conceptually near and distant communication into two, to account for the different synchronicities in phonic and graphic communication (as illustrated in Figure 2.4). Synchronicity refers to the time intervals between responses in communication: in synchronous communication, responses are exchanged in near-real time, e.g. a telephone call; in asynchronous communication, such as email, "conversation is isolated spatially, chronologically, and contextually". Within typed DMC, true synchronicity of communication cannot be achieved, as there are gaps between turns, and conversational partners do not produce simultaneous talk (Dürscheid 2003: 11). Instead, Dürscheid refers to communication

through messaging clients as 'quasi-synchronous' (for more on synchronicity in DMC, see Section 2.1.2, Anderson, Beard, & Walther 2010: 5; and Dürscheid 2003: 3-11). However, Androutsopoulos (2007: 89) criticised Dürscheid's revision, arguing it did not make an "adequate distinction [...] between interactivity in the technical sense and interactivity as a human activity"; for example, any email exchange may be asynchronous but can function quasi-synchronously due to technological advancements. Storrer (2013: 354) took a similar view to Androutsopoulos (2007) and suggested that the use of "characteristic stylistic features are not tied to the medium [...] or a particular social network" and that "writers adapt their writing style to the respective communicative setting and the appropriate linguistic conventions".

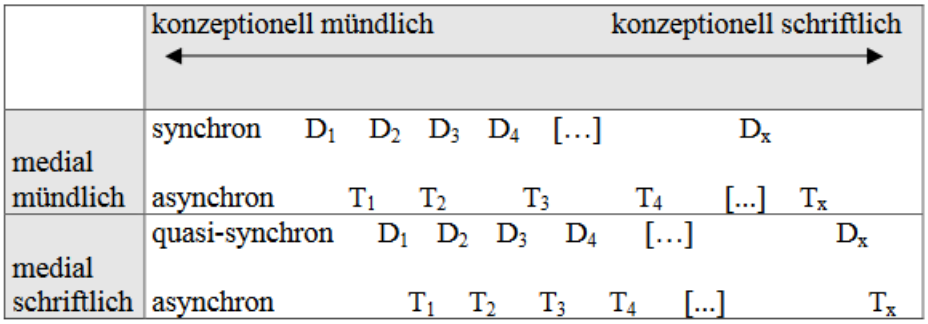



Figure 2.4: Expansion of Koch and Oesterreicher's Proximity-Distance Model incorporating synchronicity of medium (Dürscheid 2003: 13)

Androutsopoulos's criticism is part of a growing body of scholarship that has acknowledged that the *Nähe-Distanz* model is just one possible framework through which to analyse certain features of DMC. Androutsopoulos (2007) championed the use of the terms "new *Schriftlichkeit*", first coined by Haase et al. (1997: 81), and later "digitale Schriftlichkeit" ('digital literality'), as DMC was then no longer considered 'new' (Androutsopoulos & Busch 2020). Both terms acknowledge the differences between digital and traditional forms of writing, but reframe the analysis of digital features so that it is no longer constrained by the *Nähe-Distanz* model, which, as Dürscheid (2016b: 386) noted, "was never designed for this purpose [the analysis of Digital Writing]". Dürscheid (2016b: 386) went further, arguing that the "new communicative forms, particularly chat" made it almost impossible to integrate DMC, particularly Digital Writing, "within the continuum of *Nähe* and *Distanz*". Like Dürscheid & Brommer (2016),

Androutsopoulos (2007), and later Saxalber & Micheluzzi (2018), concurred that the linguistic features of Digital Writing cannot simply be treated as "a medial transposition of the aspects of spoken language" (Androutsopoulos 2007: 81). This standpoint had been prevalent in much of the older research on German, viewing Digital Writing as "emulated" spoken language (see Siever, Schlobinski, & Runkehl 2005: 7). Furthermore, even if there are linguistic themes and features that are common to much of Digital Writing, the number of online platforms and opportunities to communicate means that there are significant linguistic differences in the DMC on these platforms, as well as even within the same platform, and these linguistic features continue to evolve, diverge, and converge (and thus require constant research) (Androutsopoulos 2011). The *Nähe-Distanz* model is still useful to classify the use of certain features of Digital Writing, but does not offer a holistic solution. Instead, other frameworks are used in this study as the basis to analyse certain emergent linguistic features within Digital Writing, and are discussed below.

One of the earliest frameworks was developed to analyse interactive aspects of DMC. Collot & Belmore (1996: 15–18) applied the multidimensional-multifeature model (MD-MF), developed by Biber (1988), to DMC, to examine discourse and interaction in digital "speech situations". However, the majority of the features of digital writing identified in previous research have been examined within the framework of conceptual orality. These include interjections and terms of address (e.g. *hey*), modal particles (e.g. *doch*, *schon*), and hedging and filler words (e.g. *äh*) (see Storrer 2001; Bader 2002), as well as dialect words and slang, which Schlobinski (2005) has argued would not normally appear in written language. However, the acceptability of certain lexis in written formal language, such as in newspaper articles, is negotiable, and labels such as slang, "informal register" (Durrell 2003: 29), and 'dialect' are open to interpretation. Switzerland and Austria, for example, "each have nationally recognized and codified standards of linguistic norms" that differ from Germany (Stevenson 1995: 143), and sometimes there is a lack of an accepted standard in these countries, as in the case of Swiss-German orthography (Felder 2020). Correspondingly, on-

line language norms can also vary between the German-speaking countries.[1] Morphosyntactic features identified as indicators of conceptual orality include the use of subordinating clauses with finite verbs in the second position instead of the final position (Kilian 2001), as analysed in Chapter 5, and ellipsis, such as pronoun, article and copula deletion (Dürscheid & Brommer 2016). Orthographic and morphosyntactic features identified as markers of conceptual orality include the orthographic reproduction of phonological phenomena such as assimilation, e.g. writing < isso > to represent < ist so > with a consonant cluster reduction of < st > (see Fagan 2009: 245-248). Other conceptually oral orthographic features can include phonetic spelling (< leida > instead of < leider >), the use of 'shouting capitals' (e.g. < NEIN >), and the extended repetition of letters (< jaaaaa >) (see Kilian 2001; Androutsopoulos 2003b; Soffer 2010).

However, Androutsopoulos (2003b, 2007, 2011) disagreed with the analysis of shouting capitals and the repetition of graphemes as conceptually oral, instead developing the arguments in Sassen (2000), Schmidt (2000), and Bader (2002), to treat them as "compensation practices". These differed from conceptual orality, in that they are taken to represent "intonation patterns" rather than aspects of the register of "casual spoken language" (Androutsopoulos 2011: 149; see also Yus 2011; and Herring, Stein, & Virtanen 2013). Intonation patterns are not the only feature that is compensated in Digital Writing. There has been significant scholarship on the use of emoji and emoticons to compensate for facial expressions in face-to-face communication, as emoji and emoticons, such as < 😊 > or < :) >, represent human faces (see Miyake 2007; Beck 2010; Dresner & Herring 2010; Thompson & Filik 2016; Hougaard & Rathje 2018; Beißwenger & Pappert 2019b; Hilte, Vandekerckhove, & Daelemans 2019; Kavanagh 2019). However, the functions of emoji and emoticons are a subject of significant debate within DMC scholarship. For Albert (2015: 3), "emoticons in written language today cannot be described as a compensation strategy", rather they "have evolved into abstract, symbolic signs", which are used to modify how a message should be received. Similarly, Herring, Stein, & Virtanen (2013: 8) ar-

[1] For examples of the use of Swiss and Austrian German online, see Aschwanden (2001) (Swiss), Siebenhaar (2006c) (Swiss and Austrian), Felder (2020) (Swiss), and Glaznieks & Frey (2020) (Austrian).

gue that emoji and emoticons are not just compensatory, but that they can be used ludically, as part of "language play", which I show is indeed the case in my development of a unifying framework of digital orthographic features in Part III.

In addition to the notions of conceptual orality and compensation, Androutsopoulos (2007: 81–83) identified two further approaches to analysing linguistic features within Digital Writing: "graphostylistics", building on previous work by Sassen (2000) and Schlobinski (2001); and "language economisation" (see the overview by Siever 2006). Graphostylistics, also referred to as "graphic variation" (Spitzmüller 2013) or "graphomatic microvariation" (Dürscheid 2016a), is defined as the "manipulation of visually represented language without correspondence to phonics", i.e. stylising orthography in a visual and often ludic way (Androutsopoulos 2007: 83). This can include phonetic spellings, e.g. < kul > for < cool > (Dürscheid 2016a: 496), alternating upper and lower case letters (e.g. < kOmplImenT >, for *kompliment*, 'compliment') in a way that supposedly represents how a computer "hacker" would write; and grapheme substitution, e.g. < n8 > to represent < Nacht > and < cu > for < see you > (Androutsopoulos 2007: 84). Such stylisation has been used by commenters as an orthographic strategy to impart aspects of their identity within their Digital Writing. However, how these practices fit into wider orthographic creativity has often been overlooked within research. Accordingly, they too are integrated into the new comprehensive framework presented in Chapter 7, Section 7.4.

Stylisations, such as < n8 > and < cu >, are also often classified as "Sprachökonomie" ('language economisation'), the use of orthographic strategies for "shortening a message form" to compensate for technological and financial barriers (Androutsopoulos 2011: 149; also see Ferrara, Brunner, & Whittemore 1991: 19; Schlobinski 2006b; Siever 2006). In Digital Writing, scholarship has classified ellipsis and the use of phonetic and colloquial spellings as language economisation strategies (e.g. Siever, Schlobinski, & Runkehl 2005; Dürscheid 2005; Wirth 2005; Crystal 2008; Kessler 2008), where certain linguistic features are deliberately left out or abbreviated to save time, space, and sometimes money (in the case of SMS exchange due to users being

charged per 160 character text). With the advent of smartphones and the accompanying cheaper data packages, SMS exchange in Germany has fallen dramatically from a high of 60 billion texts sent in 2012 to 9 billion in 2018 (Statista 2021), replaced by the use of messaging clients such as *WhatsApp* and *Facebook Messenger*. Correspondingly, the removal of the financial and spatial pressures to fit a message into as few characters as possible has largely disappeared, although some features of language economisation, as well as other non-standard orthographic practices, still remain in more recent communication channels as stylistic choices, e.g. the phonetic respelling of < u > for < you > (Rotne 2018: 900), which Berg (2020) refers to as a "democratisation of orthography". However, children's use of such forms has often led to discussions over whether digital communication is negatively impacting their writing ability, rather than being a conscious choice (see Kleinberger Günther & Spiegel 2006; Androutsopoulos & Busch 2021). This area of discussion is further explored in Section 2.2 below.

One area of Digital Writing that has seen far less research is the study of the metacommunicative functions imparted by orthographic choices. The concept of metacommunication was initially developed and refined by Ruesch & Bateson (1951) in their analysis of primate behaviour as "communication about communication" that enables primates to differentiate between playful and serious versions of the same communicative act, such as between play-fighting and actual aggression (Ruesch & Bateson 1951: 88). In primates, metacommunication can take the form of "mood-signs" (Bateson 1972), such as body language or modulations in noises, which accompany the act. The concept was subsequently applied to human communication by Goffman (1974: 210), who defined human metacommunication as "a stream of signs which is itself excluded from the content of the activity but which serves as a means of regulating it, bounding, articulating, and qualifying its various components and phases". In a later paper, Goffman (1981: 280–82) defines these signs to include alteration in pitch, enunciation, speed, and volume, which alter the meaning or "footing" of the utterance.

These signs, alongside actions such as laughing and crying, are also referred

to as "paralanguage", defined by Poyatos (1993: 6) as "nonverbal voice qualities, voice modifiers, and independent utterances". However, Wharton (2016: 69–70), rejects a definition restricted to the "*vocal* aspects of language use that are not […] part of language" (emphasis in original), instead suggesting that "all those aspects of linguistic communication that are not part of language", such as gesture, facial expression, and other body language, should fall under the definition of paralanguage. In the field of DMC, the terms 'paralanguage' and 'paralinguistic' have been applied to features like shouting capitals and emoji (e.g. Merchant 2001; Kozinets 2010; Jones, Schieffelin, & Smith 2011; Pavalanathan & Eisenstein 2016), but only in regard to how they imitate face-to-face communication. Orthographic features are a part of a complex repertoire of semiotic resources of digital communication that goes beyond the mere imitation of spoken language, the aspect foregrounded in most research on 'paralanguage' in DMC to date. Therefore, the term 'metacommunication' is a preferable term to use, as it encompasses features and functions in digital writing used to communicate about communication.[2]

Androutsopoulos's framework of orthography, as detailed above, has proved useful when holistically analysing digital orthography, but it requires adaptation in order to be applied to analyse the metacommunicative functions of orthographic features in digital writing. As part of this adaptation, the metacommunicative framework developed in Chapter 7 has incorporated approaches to metacommunication that have until now solely been applied to individual groups of orthographic features, for example, the approach to "digital punctuation" by young German speakers, which investigated how punctuation marks can have metacommunicative functions (Rinas & Uhrová 2016; Androutsopoulos & Busch 2021), including even indexing certain identities (Androutsopoulos 2018, 2020b). In addition to punctuation, an increasing body of research, particularly since 2015, has examined the functions of emoji in digital writing (e.g. Pavalanathan & Eisenstein 2015; Ljubešić & Fišer 2016; Evans 2017; Beißwenger & Pappert 2019a; Dainas & Herring 2020), as well as related graphical features

[2]The term 'metacommunication' has been also used, in a different context, to refer to conversations that explicitly refer to previous conversations (Jones, Schieffelin, & Smith 2011; Castor 2017: 4), but that is unrelated to the sense in which it is used here.

that are specific to other platforms such as personalised bitmoji on *SnapChat* (Danesi 2016: 60-61), as shown in Figure 2.5; the augmented-reality animoji on iOS (Herring et al. 2020a; Herring et al. 2020b), as shown in Figure 2.6; and emotes on *Twitch* (Barbieri et al. 2017). While there is some overlap with early German-language DMC research regarding the use of emoji to compensate for a lack of face-to-face communication (e.g. Novak et al. 2015; Seargeant 2019: 6), several of the functions regarding emoji use identified in previous research also discuss their potential applications in the wider digital communication environment, such as the ludic use of emoji identified by Kelly & Watts (2015), or the use of emoji to structure a message (Dürscheid & Siever 2017), which has similarities with the creative use of spacing, examined in Section 7.1.2.

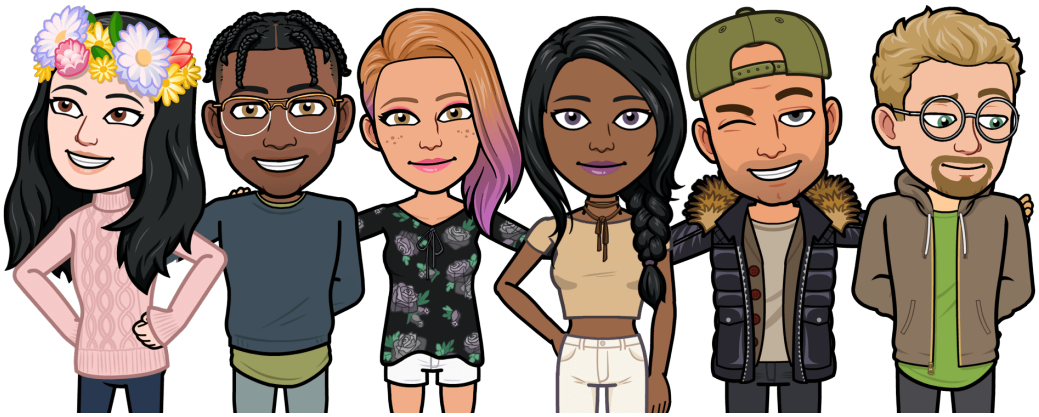


Figure 2.5: Examples of Bitmoji, from Snapchat Inc. (2021)

Figure 2.6: Three states of the 'Unicorn' animoji

In addition to studies of orthographic features, I draw on relevance theory (see Wilson & Sperber 2002) to develop the framework of metacommunication presented here. Relevance theory seeks to fully understand the Gricean maxim of relation (Grice 1975), i.e. the "expression and recognition of intentions [by a speaker] on the basis of the evidence provided" (Wilson & Sperber 2002: 606), the evidence including the sum of all linguistic, metalinguistic, and sociolinguistic aspects, such as lexical and orthographic choices, and their interactions with the interlocutors' lived experiences. For example, while the various smiling emoji used in Example 2.1 can be classified as a compensatory measure to represent the happy demeanor of the author, the use of the assorted emoji at the end of the comment fulfils different metacommunicative functions, discussed below, which are a product of the available linguistic resources in DMC, and not just a compensation strategy for the lack of face-to-face communication.

**Example 2.1** (2017)
Frühstück am Bett.... Luxus 😇😂😄😘 nices viedeo 🐴🎁📻📷💻
('Breakfast in bed.... luxury 😇😂😄😘 nice video 🐴🎁📻📷💻')

The full spectrum of metacommunicative functions within DMC has, I argue, never been limited to mere representations of spoken language or compensation for the lack of features available in face-to-face communication. This recognition requires the development of a more substantial framework, which focuses on the shared semiotic resources that are created and developed using the available linguistic resources on communication devices (in this case: the inputs allowed by a mobile phone or computer) in online spaces (in this case: the *YouTube* comments section only allowing textual communication). Furthermore, youth language has often not explicitly been used as the data source in previous scholarship for such investigations, despite the high proportion of young people who regularly communicate digitally (Bahlo et al. 2019: 80). The lack of a holistic approach to metacommunication is addressed in the analysis of online youth language presented in Part III. The analysis presented in these chapters also incorporates research on the expression of identity and interaction through language, which falls under Digitally Mediated Discourse. An

overview of research in this field is given in Section 2.1.2.

### 2.1.2 Digitally Mediated Discourse

While scholarship on Digital Writing focuses on the linguistic features that are used, work on Digitally Mediated Discourse concerns the contexts of language use. In this section, I focus in particular on research on conversation and interaction, and on the performance of social identities (see Butler 2006), the two aspects most pertinent in analysing *YouTube* comments written by young people.

DMC research on interaction has often used approaches from Conversation Analysis (Schegloff & Sacks 1973; Sacks, Schegloff, & Jefferson 1974; Sacks & Jefferson 1992; Hutchby & Wooffitt 1998; ten Have 2007; Liddicoat 2022), focusing, for example, on turn-taking (the alternation of who next "talks," see Herring 1999; Riva 2002; Kessler 2008; Bou-Franch, Lorenzo-Dus, & Blitvich 2012; Androutsopoulos & Tereick 2016; Meredith 2019); openings and closings (starting and finishing conversations, see Kessler 2008; Meredith 2019); and topic structure and shift (e.g. the change in conversational subject, see Herring 1999, 2013; Dowell et al. 2017). From very early on in DMC scholarship, it was acknowledged that different modes of DMC (e.g. mail, chat) produced not just different linguistic features, but also different interactive styles, as noted in Werry (1996), often related to the synchronicity of the mode of communication (Frehner 2008: 168). More synchronous communication is characteristically more dialogical with rapid alternation of turns (Crystal 2008), and contains more topic shift (Herring 1999, 2013), as well as lexical features that represent openings and closings (e.g. "hi") (Kessler 2008). Siever, Schlobinski, & Runkehl (2005), Wirth (2005), and Kessler (2008) have suggested that more synchronous situations encourage language economisation and conceptual orality, in both German and English-language DMC.

However, there is also evidence that the link between the synchronicity of a DMC mode and the choice of certain linguistic and interactive features is not this straightforward. Dürscheid (2005) and Gibson (2008) both demonstrated

that openings and closings, characteristic of synchronous communication, were also widely used in both email and Virtual Learning Environments (VLEs), which had been considered asynchronous. Similarly, Androutsopoulos (2015) demonstrated that communication between participants on *Facebook* (on a user's profile page, or 'wall') can exhibit qualities of both asynchronous and synchronous communication, in that posts may be responded to almost immediately, after several hours or after days, yet the comments, regardless of the time gap, contain examples of features characteristic of synchronous communication, such as ellipsis. The use of features both categorised as synchronous and asynchronous is also evident within *YouTube* comment threads (Bou-Franch, Lorenzo-Dus, & Blitvich 2012; Androutsopoulos & Tereick 2016). Spaces such as *Facebook* and *YouTube* comments sections were "not designed as a Web space for collaboration and synchronous interaction" (Chau 2010: 72), yet they often exhibit hybrid interactive practices, "characterized by a combination of, on the one hand, the orderly, turn-by-turn patterns typical of dyadic conversation and, on the other hand, 'networked sequences' consisting of adjacent and nonadjacent turns typical of asynchronous interaction" (Bou-Franch, Lorenzo-Dus, & Blitvich 2012: 503). These hybrid practices can be seen in many examples from the *NottDeuYTSch* corpus presented throughout this study, particularly revealing how such practices influence the use of lexical items as different parts of speech (Chapter 4), as well as potential grammaticalization processes (Chapter 5). However, developments in syntax are rarely addressed in DMC scholarship on the German language, despite the large amount of Digital Writing in corpora available for analysis, such as the *DWDS WebXL Korpus*. The structure of the interaction between commenters – not addressed in detail in this study – is also a topic ripe for future research (see Cotgrove, Conklin, & Thul, in prep.).

In addition to research on the structural aspects of DMC interaction, a significant proportion of scholarship also deals with social aspects of interaction features, especially concerning user identity (or anonymity) and the construction of online communities (for a discussion of anonymity, pseudonymity, and online identity, see Döring 2010). Herring (2019: 31–32) notes that the first online communities were interest-based and this is discussed in early scholarship,

which examined, for example, newsgroups for political discussion (Gruber 1997; Jones 1998; Papacharissi 2004), mailing lists for hobbies and interests (Erickson 1999; Bell & Hübler 2001; Dresner & Herring 2010), and MUDs for role-playing games (Danet 1998; Kendall 1998; Utz 2000; Nakamura 2002).

Within this early research there was the hope that online anonymity might lead to a socially equal space (Herring 1996b), but in fact it often led to "uninhibited verbal behavior [...] swearing, insults, name calling, and hostile comments" (Kiesler, Siegel, & McGuire 1984: 1129). This was found to be symptomatic of masculine posturing (Jones 1998: 59). This "pre-web (1983-1993)" period of DMC was mostly populated by white men from the USA and the UK, although from the early 1990s, there was an "increase in female users" (Herring 2019: 39). This perceived increase led to the socially gendered aspects of DMC becoming a prominent topic in scholarship during the 1990s, and it has continued to be an important topic, broadly applying developments in 'offline' sociolinguistic research to discourses in online communication.[3]

Prominent research themes in the 1990s focused on potential differences in linguistic practices between men and women, for example, a possible tendency for men to use more assertive language, swearing, and sarcasm, and for women to use cooperative language, hedges, apologies, and questions (e.g. Herring 1992, 1996c, 1996b; Savicki, Lingenfelter, & Kelley 1996; Schwartz et al. 2013). Other work has considered possible gendered differences in the usage of DMC-specific linguistic features. For example, it was suggested that the use of emoticons, such as < :) >, and punctuation marks were a characteristic of women's language in DMC (Witmer & Katzman 1997; Wolf 2000; Baron 2004; Waseleski 2006; Parkins 2012; Schwartz et al. 2013). However, Huffaker & Calvert (2005), in a study of blogs, found no gender-based differences in lexical choice, and also that (young) men used more emoticons than women. Furthermore, Hilte, Daelemans, & Vandekerckhove (2020) suggested that age may be a more important factor in determining how emotionally expressive a DMC text may be,

[3] Space precludes full discussion of the considerable body of work on social gender and language that has been referenced in DMC scholarship, but notable studies include Lakoff (1975), Cameron (1992), and Holmes & Meyerhoff (2003) in an English-language context, and Trömel-Plötz (1979), Trömel-Plötz (1982), and Hellinger (1995) in a German-language context.

with younger people's language containing more expressive features.

The availability of larger amounts of DMC data since the mid-2010s has enabled more nuanced approaches to analysing emoticons and emoji (and other linguistic features). For example, Fladrich & Imo (2020) used the *MoCoDa2* corpus of German-language *WhatsApp* conversations to analyse the use of particular emoji as an indicator of gender identity, finding significant differences between the top twenty emoji used by women and men in both mixed and single-gender settings. A burgeoning body of research has taken larger-scale computational linguistic approaches to emoji use with the aim of identifying the gender identity of the user (e.g. Chen et al. 2018; Jaeger et al. 2018; Koch, Romero, & Stachl 2020). It should be noted that analyses largely still operate within the framework of a binary notion of gender, possibly reflecting the lack of implementation of technical means in the data sources, in these cases, Twitter or *WhatsApp*, for users to express non-binary gender identity.[4] The quantitative and qualitative methodologies used in Part III to analyse orthographic features, particularly the features used to perform identity, are informed by the approaches taken by the computational analyses mentioned above, but are not restricted to any binary notions of gender.

Following the rise of DMC research on gender, other social aspects of identity began to be analysed from the late 2000s onwards, such as how language is used to construct and maintain communities, as well as sexual, regional, and ethnic identities. Such online communities have been referred to as "speech communities" in the sense of Gumperz (2009: 66), i.e. "any human aggregate characterized by regular and frequent interaction by means of a shared body of verbal signs". Gruzd, Wellman, & Takhteyev (2011) preferred the term "imagined communities" (from Anderson 1983), understood as a group of people with shared interests or a shared identity who do not necessarily interact with each other. Varis & van Nuenen (2017: 478) observed that online communities do not require "temporal and spatial co-presence", a fact which "force[s] us to reconsider established understandings of 'community' "; they instead preferred

[4]The *MoCoDa2* corpus does include a *divers* ('diverse') gender option, see Beißwenger et al. (2020), the official terminology that serves as a catch-all for gender identities not covered by the traditional binary.

to describe online communities as "translocal".

There was also a shift in the late 2000s from research on generalised linguistic variation, such as binary gender language differences, to analyses of 'doing' identities in online situations. The concept of 'doing' identity stems from research on gender by West & Zimmerman (1987: 125), which posits that gender is "a routine accomplishment embedded in every interaction", i.e. it is not an immutable property of existence, rather it is socially constructed and "performed" (Butler 2006: 187). This understanding of gender has been applied to other social identities, such as ethnicity ('acts of identity,' Le Page & Tabouret-Keller 1985; 'ethnifiying,' Lytra 2016) and youth ('doing youth,' Neuland 2003; Walther 2018).

In sociolinguistics, this approach to language and identity falls under what Eckert (2012) has called "Third Wave Variationist" sociolinguistics, i.e. examining how language variation is used to construct meaning, identity, and style, which are inherently "mutable" (Eckert 2012: 94). For example, in DMC scholarship, Blashki & Nichol (2005), Milani & Jonsson (2011), and Heritage & Koller (2020) analysed linguistic features of heterosexual masculinities in online men's communities, such as the use of language to create a 'geek' identity, or to promote discourses of sexism and misogyny. Dmitrow-Devold (2017) examined the gendered performances of teenaged girls in blogs, Mackenzie (2018) investigated the performance of motherhood in online forums, and Willem, Araüna, & Tortajada (2019) analysed sexist and classist language expressing sexualised stereotypes of women. A similar focus on identity is found in some studies of the language in *YouTube* comments. Bou-Franch, Lorenzo-Dus, & Blitvich (2012), Bou-Franch & Blitvich (2014), Colliver & Coyle (2020) have all examined discourses of violence against women in *YouTube* comments. Wotanis & McMillan (2014), Potts (2015), Abidin (2019), Döring & Mohseni (2019), Haryanto & Suwito (2020) examined attitudes towards gender and sexuality identities perceived to challenge heterosexual masculine 'norms'.

Other areas of Digitally Mediated Discourse on *YouTube* have focused on the use of language to perform ethnicity, e.g. the use of humour to construct Arab

and East Asian identities (Chun & Walters 2011), the use of stereotypically 'Black' language as part of Asian-American identities (Chun 2013), and the importance of language to Ukrainian (Tovares 2019), Eastern Maroon (Migge 2020), and Quechua identities (Cueva et al. 2021). Other linguistic scholarship on ethnicity on *YouTube* has examined identity work in song lyrics (e.g. Mendoza-Denton 2016; Cotgrove 2018; Røyneland 2018; Garley 2019) and multilingualism in *YouTube* comments (Thorne & Ivković 2015). Multilingualism has also received significant attention in other DMC scholarship on language and ethnicity, for example on online diasporic websites (e.g. Lo 1999; Androutsopoulos 2006; Paolillo 2011; Wiese 2015; Hinrichs 2018), e.g. "codeswitching" between the language of the country of residence and 'heritage' languages, or, more recently, on examining the use of multiple linguistic resources within the same communicative act as "translanguaging" (García & Li 2014; see also alternative, overlapping terminology, e.g. 'codemeshing,' Canagarajah 2011; 'metrolingualism,' Pennycook & Otsuji 2015). It is important to note that multilingualism is common throughout the *NottDeuYTSch* corpus. Anglicisms, in particular, are often used in German youth language, as we shall see in Section 2.2, a fact which informed the investigation of lexis presented in Chapter 4. There are also possible instances of multilingual influence on syntax in the *NottDeuYTSch* corpus, as discussed in Chapter 5, a topic that is rarely covered in DMC scholarship (for spoken language, see Wiese 2009; Wiese, Öncü, & Bracker 2017).

DMC scholarship on identity and community has not been restricted simply to gender, sexuality, and ethnicity. First, within German-language research, there has also been a long history of scholarship on the construction of regional identity in DMC, partly due to German's status as a pluricentric (and plurinational) language, i.e. a language with multiple "varieties of the standard across different regions" (Ammon 2005: 31). For example, Siebenhaar (2006a), Siebenhaar (2006c), and Durham (2007) studied Swiss German varieties within chatrooms and mailing lists; Saxalber & Micheluzzi (2018) investigated the linguistic construction of South Tyrolean identities on *Facebook*; and Androutsopoulos (2013) analysed data on discourses of Berlin and Bavarian identities in *YouTube* comments. Second, as shared interests remain an important reason for inter-

action on the internet, considerable scholarship has examined how language shapes identity and community in particular online spaces. For example, Androutsopoulos (2003b) examined speech variation across several sites of DMC, including forums, IRC, and online guestbooks, and there has been sustained research on language and communities on the microblogging website, Twitter (e.g. Marwick & boyd 2011; Gruzd, Wellman, & Takhteyev 2011; Papacharissi 2012; A Scott 2015; Panizo-LLedot et al. 2019; Ilbury 2020), as well as the online content aggregator, *Reddit* (LaViolette 2017; Flesch 2019; Heritage & Koller 2020).

Scholarship specifically on *YouTube* has often investigated metalinguistic attitudes towards the linguistic expression of identities in the comments sections, i.e. how commenters view the use of particular linguistic features as part of the identities that are being performed. Examples of previous studies on the metalinguistic discussion of identities in *YouTube* comments have investigated attitudes to linguistic markers of class identities (Aslan & Vásquez 2018), Northern Irish (Lee 2016), and multiple US-American accents (Rymes & Leone-Pizzighella 2018; Cutler 2020). Similarly, Cutler (2016) analysed comments under a *YouTube* video about the animated film *Brave* to demonstrate how non-standard orthography can be used to represent Scottish accents. In a German-language context, scholarship has focused on how the use of non-standard orthography in *YouTube* comments is key to metalinguistic discourse on dialect. For example, Androutsopoulos (2010, 2013) and Androutsopoulos & Tereick (2016) investigated the use of markers of several German dialects, including Bavarian and Berlinerish, and the responses to and evaluations of such use in comments sections.

Scholarship on language and identity has historically tended to foreground one particular aspect of the subjects' identities, such as gender, ethnicity, class, or sexuality. Levon (2015: 295) critiqued the "inadequacy of isolated categories in sociolinguistic analysis", calling for a more intersectional approach (see Crenshaw 1989; Hill Collins 2002; Yuval-Davis 2011). An intersectional approach is therefore taken throughout this thesis, most prominently in Chapter 7, in an analysis of how the intertwined aspects of identity can affect orthographic

choice. However, one aspect of identity that has not received as much prominence in DMC research is youth, which is central to this study. I turn in Section 2.3 to a detailed review of scholarship on the intersection of DMC and youth, but first, present an analysis of the work to date on German youth language in Section 2.2.

## 2.2 The language of young people

The centrality of language for youth identities cannot be overstated. As a "reaction to general conditions of adolescence as a transitional life stage", it is "a symbolic assertion of autonomy and [...] an index of affiliation to (or distancing from) relevant peer groups and youth-cultural scenes", as well as a "a departure from mainstream norms and dominant cultural values" (Androutsopoulos & Georgakopoulou 2003: 4; see also Dürscheid, Wagner, & Brommer 2010; Tagliamonte 2016a, 2016b). The friction between 'youth' and 'adulthood' leads to linguistic innovation, and the present study demonstrates such innovation in the areas of lexis, morphology, syntax, and orthography in the *NottDeuYTSch* corpus. The analysis presented here is informed by work since the 1980s, which saw renewed attention to contemporary linguistic aspects of German youth language (Androutsopoulos 2008). Notable studies from this early period include Henne (1986), who constructed an overview of the lexis and phraseology of young people by documenting hundreds of words and phrases from secondary school pupils from Braunschweig and Mannheim in 1982; and Neuland (1986), who examined the contexts of German youth language in times of social change, as well as Heinemann (1979) and Funke (1982), who investigated youth language in the GDR. Neuland (1987: 59) observed that academic interest in youth language stemmed from discourse surrounding the "youth revolts" at the end of the 1970s. The hostility of many youth 'scenes' at the time to institutional norms caused methodological difficulties, particularly regarding the collection of data, as researchers, as 'outsiders' to the scenes, were shut out and unable to observe the groups.

During the 1990s and early 2000s, scholarship expanded through the analysis of

written sources of *Jugendsprache*, such as magazines (both professionally published and fan-made 'zines') and song lyrics (e.g. Schlobinski, Kohl, & Ludewigt 1993; Schlobinski 1995; Neuland 1994; Androutsopoulos 1998). Androutsopoulos (1998: 1–2) described the field of *Jugendsprache* as split into two "camps" of academic thought based on the focus of the research: speaker-oriented (focused on the 'user' of the language), or system-oriented, i.e. focused on social structures and systematic aspects of language, referred to as 'uses' in Holmes (2013: vii). In the overview below, I combine the terminology and approaches used by Androutsopoulos (1998) and Holmes (2013) to examine previous scholarship on *Jugendsprache* under the categories of 'user-oriented scholarship' and 'system-oriented scholarship'.

### 2.2.1 User-oriented approaches to youth language

Early user-oriented research argued that youth language is inherently experimental, particularly towards conventions of language, and disagreed with the contemporary portrayal of German youth language as simply specific lexis or expressions (see Schlobinski, Kohl, & Ludewigt 1993; Neuland 1994). Several early user-oriented studies characterised youth language as a "neutral" variety of German with its own linguistic features (Schlobinski 1995: 320), to combat its previous depiction in media outlets and among older academics, as a homogenous "brutal", "vulgar", and "rowdy language" (Schlobinski, Kohl, & Ludewigt 1993: 9-10), apparently responsible for the impending downfall of the German language (see Neuland 1994: 81; Schlobinski 1995: 317; Eichinger 2001). User-oriented scholarship has often taken ethnographical approaches to linguistic data, focusing on how young people "construct their identity" through "specific ways of speaking and writing" (Henne 1981: 373).

Such identities are manifold, and a central strand of early ethnographical research acknowledged the plurality of youth groups and scenes and the correspondingly different speech styles used within those communities. Schlobinski (1995: 336) rejected the notion of *Jugendsprache* as "a monolithic concept", stating that "there are as many *Jugendsprachen* as there are groups" (Schlobinski

1995: 326). This focus on 'scenes' and groups was continued by Androutsopoulos (1998) and Watzlawick (2006: 290), who identified lexical and morphosyntactic differences between 'subcultures', including heavy-metal fans, skinheads, and Hip-Hoppers. During the 2000s, ethnographical research on youth languages focused less on 'scenes', and more on larger social identities, such as minoritised ethnicity and gender within groups of informants, often taking an intersectional approach. Most notably, community languages are the subject of much scholarship, in particular Turkish, in the language of young German speakers, referred to variously as *Lansprache* (*lan*: Turkish for "man; dude" + *Sprache*: German for "language," Kallmeyer & Keim 2003), *Türkenslang* ("Turk-slang," Auer 2003), and *Kiezdeutsch* ("hood-German," Wiese 2009, 2012, 2022; Wiese & Rehbein 2016). Auer & Dirim (2003) and Deppermann (2007) examined the acquisition of Turkish words and phrases by young people without Turkish heritage (for a similar phenomenon in the UK, see Rampton 2014). Kallmeyer & Keim (2003), Keim & Knöbl (2007), Keim (2007b), and Keim (2007a) investigated the language of multicultural and Turkish-German friendship groups of girls in Mannheim. Similarly, Bierbach & Birken-Silverman (2007) examined gender differences in language choices within a group of young German-Sicilians in Germany and Spreckels (2006) investigated the linguistic construction of a gendered "in-group" and "out-group" (see Padilla 1999: 116) among young German women without a 'migration background'.

Despite the focus on linguistic differences between groups of young people, there are also considerable overlaps that reflect the shared aspect of a youth identity. For example, most informants in Watzlawick (2006) gave *geil* ('cool') as their preferred word choice for positive evaluation, and *Penner* ('tramp') for a negative word for a person. The shared aspects of youth identities have informed Chapter 4, which analyses a range of words used for positive evaluation spanning multiple scenes, and demonstrates their high productiveness but also propensity to change in frequency over a relatively short space of time. Within scholarship to date, there have been few studies that have examined syntax that is characteristic of *Jugendsprache*, notable examples being the influence of Turkish on German syntax in Wiese (2009) and Wiese (2012). Indeed, most syn-

tactic features of *Jugendsprache* are often conflated with spoken language (see Bahlo et al. 2019: 147). In Part II, Chapters 5 and 6 present a comparative study of several corpora to challenge this assumption, and also analyse the increase in the use of certain syntactic phenomena amongst young German speakers.

### 2.2.2 System-oriented approaches to youth language

While ethnographic scholarship focused on the differences in language use between groups of young people, research on the linguistic features and frameworks of German youth language sought to identify common characteristics and structures that can be generalised as representative of youth language, whilst acknowledging that youth language was a heterogenous collection of language styles. Such system-oriented scholarship initially characterised German youth language as an "age-specific variety of informal communication", with mostly "lexical-semantic" particularities (see Henne 1981; Beneke 1985; von Polenz 1990), but soon expanded, identifying as characteristic the use of certain interjections (*ej/ey*, see Henne 1986; Schlobinski, Kohl, & Ludewigt 1993), specific words corresponding to value judgements (*geil*, 'cool', see Neuland 1994; Schlobinski 1995), references to people or groups (*Alter*, 'man/dude'), and a vulgar or taboo lexicon (*abkacken*, 'to mess up', see Androutsopoulos 1998; Androutsopoulos & Scholz 1998). One of the largest claimed sources of *Jugendwörter* was music, most notably English-language music (Henne 1986; Schlobinski, Kohl, & Ludewigt 1993), adopting and adapting English words and phrases from the music scene into German youth language, such as *Pop-Star*, *Show*, and *Sound*. The use of Anglicisms in German youth language (and German in general, see Onysko 2007; Pfalzgraf 2009; Burkhardt 2013) has continued to generate significant research (e.g. Zifonun 2000; Androutsopoulos 2005; Wittenberg & Paul 2009; Cotgrove 2013; Bahlo & Klein 2017; Görke 2018; Mešic 2018; Neuland 2018; Preseau 2018), as have borrowings from other languages, most often community languages, as discussed above in Section 2.2.1. Although system-oriented research has since progressed beyond the collection of certain characteristic words and expressions, the image of youth language as young people's jargon continues within popular media. For example, the dic-

tionary publishing company Langenscheidt has annually published the "100% Jugendsprache" series of books (e.g. Langenscheidt 2014), and also annually compiles a list of "Jugendwörter des Jahres" ('youth words of the year') (Langenscheidt 2018), reinforcing the popular focus on lexis.

The collection of *Jugendwörter* is partially a response to the relationship between German youth language and language change (see Eichinger 2001; Neuland 2006), i.e. the potential adoption of words that were considered a part of youth language into standard German, such as *cool* (Androutsopoulos 2005), or general informal German, e.g. *geil* (see Chapter 4 of this thesis). Androutsopoulos (2005) referred to German youth language in this process as "lexikalische Erneuerungsquellen des Standards" ('sources of lexical renewal of the standard'), drawing on a six-stage model of linguistic innovation by Kotsinas (1997) that showed how linguistic innovations by young people can eventually spread into widespread usage (Figure 2.7).

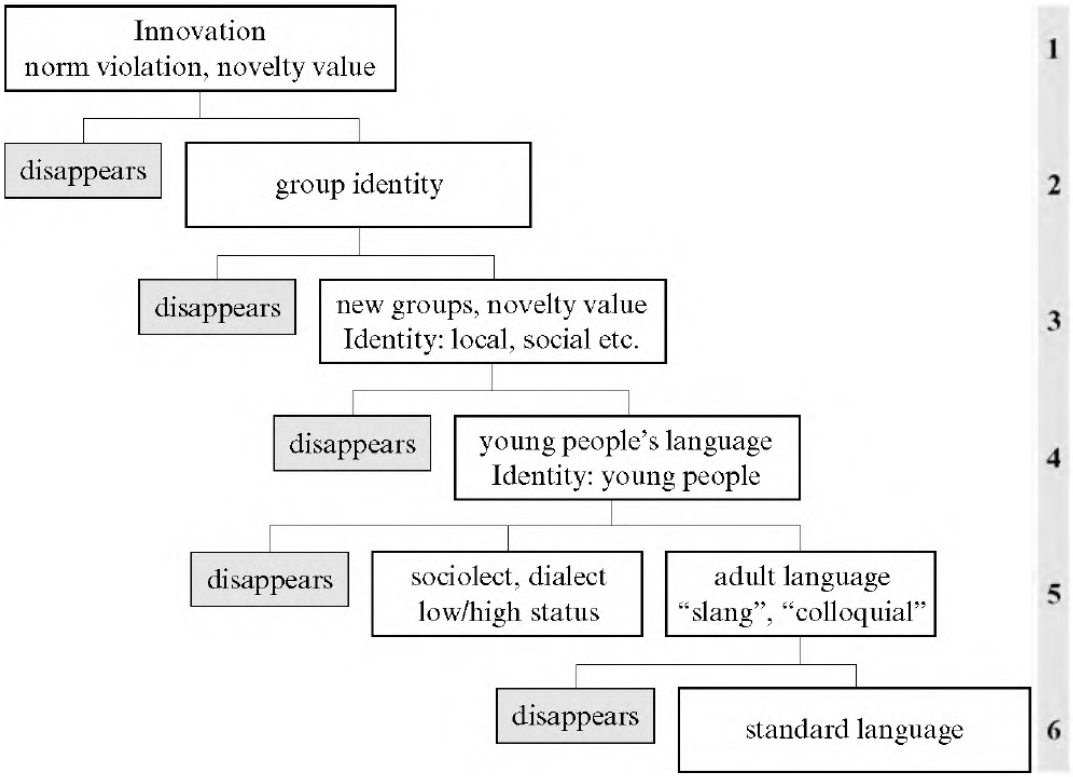



Figure 2.7: Model of linguistic innovation (Kotsinas 1997), taken from Androutsopoulos (2005: 182)

In Kotsinas' model, linguistic innovations can 'progress' up the stages as they are adopted by increasingly larger and more general groups of speakers, al-

though they are most likely to quickly fall out of use or remain in use at lower stages. The model of linguistic innovation was refined by Bahlo (2021) (see Figure 2.8), to illustrate that innovation may be cyclical, i.e. words that have been adopted into higher stages may be subject to recontextualization or further innovation by young people, thus beginning the cycle of innovation anew. The models inform the investigation in Chapters 4 and 5, alongside quantitative analyses of the *NottDeuYTSch* corpus data, to track lexical items and syntactic phenomena that have become part of the repertoire of mainstream youth culture over the timeframe of the corpus.

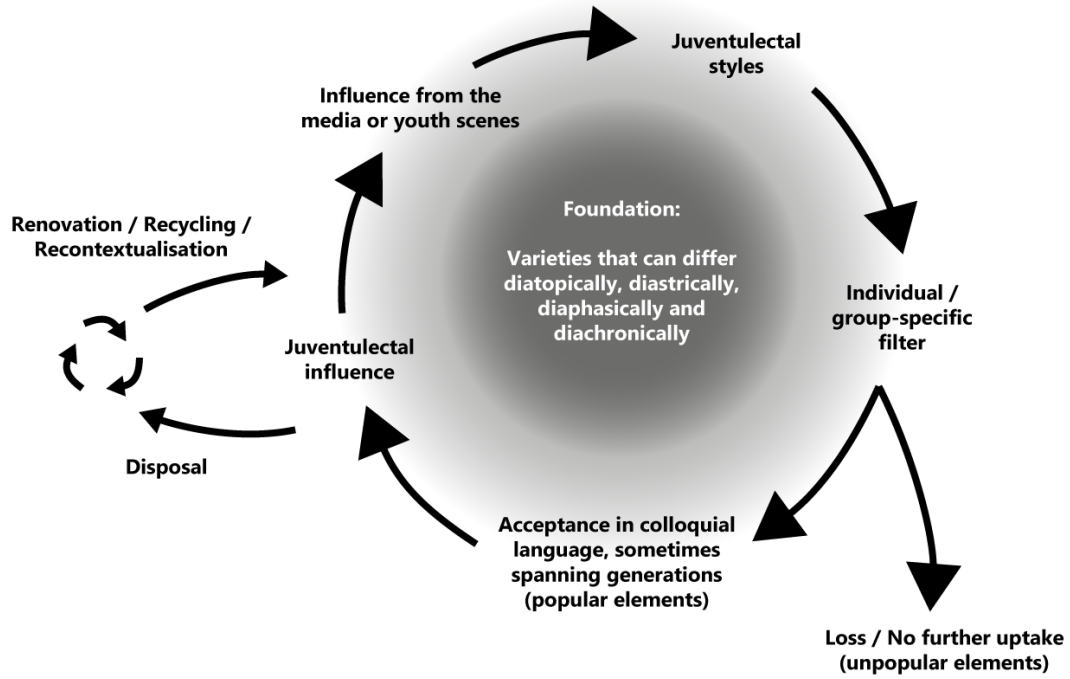



Figure 2.8: Model of linguistic innovation, adapted from Bahlo (2021a)

Schlobinski, Kohl, & Ludewigt (1993) and Neuland (1994) first identified preliminary morphological and structural features of German youth language, such as the < rum- > verb prefix (e.g. < rumsauen >, 'to make a mess'). Since then, much further scholarship has presented overviews of the linguistic and grammatical structures and features of German youth language, e.g. Androutsopoulos (1998), Pauli (2011), Bahlo & Klein (2017), Neuland (2018), Oberdorfer & Weiß (2018), Walther (2018), Bahlo et al. (2019), and Könning (2020). Synthesising these overviews yields five areas of innovation that are characteristic of

German youth language:

1. Lexical innovation (Androutsopoulos 1998; Pauli 2011).
2. Morphological productiveness (especially innovations in compounding and derivation, Bahlo et al. 2019).
3. Non-standard syntax (Bahlo & Klein 2017; Oberdorfer & Weiß 2018)
4. Orthographic variation (Androutsopoulos 2000; Bahlo et al. 2019).
5. Discursive functions (Androutsopoulos 2008: 1498).

These have directly informed the areas chosen for analysis in this study, in particular points one and two, which are discussed in Section 2.2.3, point 3, discussed in Section 2.2.4, and point 4, which is covered in Section 2.2.5.

### 2.2.3 Lexical and morphological applications

Chapter 4 builds on previous research on lexical innovation and morphological productiveness, focusing on the use of the word *geil*, which, as mentioned in Section 2.2.1, has been a key lexical item in previous scholarship on youth language. One of the earliest attestations of *geil* noted in scholarship on German youth language is reported by Henne (1986: 35–36), as an attributive adjective as part of the phrase *geile Mucke* ('cool music'), and the word has been regularly observed in studies of young people's language since the 1980s. However, in more recent years, the characterisation of *geil* as prototypically *jugendsprachlich* has been the subject of debate. While *geil* has traditionally been referred to as archetypical of youth language (Eichinger 2001: 347; Bahlo et al. 2019: 56), or "appropriate only in youth in-group communication" (Bierbach & Birken-Silverman 2007: 150), since the 1990s it has also been argued that *geil* is no longer exclusive to youth language. Several studies have argued that *geil* has been adopted into colloquial adult language, as illustrated in Chapter 2, Section 2.2.2 (e.g. Schlobinski, Kohl, & Ludewigt 1993: 63; Schlobinski 1995: 337; Deppermann 2007: 53; Hofmann 2018: 74; Androutsopoulos 2005: 175). Indeed, by 2007, *geil* was no longer marked as *jugendsprachlich* in the Duden, but rather more broadly as *umgangssprachlich* ('colloquial') (Földes 2019: 133). Despite its potential change in status, *geil* still frequently occurs in more recent

studies of youth language, which have claimed that the word remains a "core lexical item" in the repertoire of young German speakers (Vasiljevič 2018: 920), in both spoken (e.g. Henne 1986; Bierbach & Birken-Silverman 2007; Könning 2018) and online settings (e.g Kessler 2008; Voigt 2011), including in language used in *YouTube* videos (e.g. Bahlo et al. 2019: 200-201).

Scholarship has also examined potential alternatives to *geil* within German youth lanuage, acknowledging that the lexis of young people is liable to change rapidly within a relatively short space of time compared to standard language (Bahlo et al. 2019: 13-14), particularly evaluative expressions (Mroczynski 2018: 325). Mroczynski (2018) investigated whether the use of *porno* was becoming more frequent than *geil* as an emphatic interjection in German youth language, using several small corpora of spoken and online German, but found that "*geil* is still very present, both in spoken and written language", and that the emergence of *porno* was just a "flash in the pan" (Mroczynski 2018: 339). Other suggested alternative *jugendsprachlich* positive evaluations have included *cool* (Zifonun 2000; Auer 2016; Vasiljevič 2018), *easy* (Zifonun 2000), *fett* (Hofmann 2018), *genial* (Zifonun 2000; Auer 2016), *klasse* (Auer 2016), *korrekt* (Zifonun 2000; Androutsopoulos 2001), *krass* (Androutsopoulos 2001), *nice* (Könning 2018), *stark* (Androutsopoulos 1998; Auer 2016), *super* (Androutsopoulos 1998), and *toll* (Androutsopoulos 1998; Auer 2016), although none of these studies focused on whether usage of *geil* was falling in favour of any of the listed terms above.

The uncertainty surrounding the precise status and usage of *geil* also makes it an ideal subject for longitudinal analysis, and the ten-year time period covered by the *NottDeuYTSch* corpus provides the data for such an investigation. The data in the *NottDeuYTSch* corpus also provides a continuation of the longitudinal analysis of *geil* and *cool* in the *DeReKo* examined in Androutsopoulos (2005). As we shall see in Section 4.4.1, the frequency of *cool* has in fact overtaken that of *geil* in the *NottDeuYTSch* corpus.

Furthermore, there have been few studies that have explicitly investigated the specific ways in which *geil* is used in *Jugendsprache*, such as the broad range

of lexical innovations for which *geil* can be the base lexeme or affix. Androutsopoulos (1998), in a linguistic analysis of youth-produced 'zines' (low-budget magazines produced by those affiliated to a particular youth culture), provided examples of lexical, morphological, and phraseological phenomena containing *geil* characteristic of German youth language, thereby demonstrating its productiveness within the language of young people in the 1990s. The present study is informed by findings in Androutsopoulos (1998) but examines a more recent and far larger dataset to examine linguistic developments in the usage of *geil*.

### 2.2.4 Syntactic applications

Chapter 5 focuses on the use of non-standard syntax by young people, in particular, the phenomenon of main clause structure in subordinating clauses, which is referred to as parataxis (as opposed to hypotaxis). Non-standard paratactic clause structure is a hotly debated area of scholarship and has variously been called "Nicht-VL-Sätze mit Nebensatzmarkern" ('Non-verb-last-clauses with subordinating markers'), "V2-Nebensätze" ('verb-second subordinating clauses'), "eingebettete V2-Sätze" ('embedded verb-second clauses'), and "V2-Komplementsätze" ('V2 complement clause') (Freywald 2016: 329). The phenomenon has been shown in previous studies to occur primarily with clauses introduced by *weil* and *obwohl*, and usage has been reported as increasing since the 1980s (Günthner 1996: 323), although empirical data are lacking.

The word order of clauses introduced by subordinating conjunctions, particularly *weil*, began to receive significant scholarly attention in the 1980s. The research often highlighted that the use of paratactic structures in clauses introduced by a subordinating conjunction was a "recent" phenomenon in German, or at least suggested that the usage had increased noticeably in recent times (cf. Gaumann 1983: 152; Günthner 1996: 323).[5] However, Freywald (2010: 59–60) showed that there is no consensus among academics either on when

[5] In many papers, "recent times" is not clearly defined but is implied to be roughly 20 years before the paper was written, at the earliest since the 1950s.

the phenomenon took hold or when it substantially increased. There is broad agreement, however, that paratactic word order following subordinating conjunctions is a feature of spoken and/or colloquial language, rather than formal written language, and occurs following either causal conjunctions, e.g. *weil*, or concessive conjunctions, e.g. *obwohl* and *wobei* (Gaumann 1983; Günthner 1993, 1999, 2000a; Wegener 1993, 1999; Keller 1995; Uhmann 1998; Miyashita 2003; Schwitalla 2003: 144; Freywald 2009, 2010, 2016; Antomo & Steinbach 2010; Breindl 2012; Reif 2012; Reis 2013; Ágel 2016; Kempen & Harbusch 2016; Schäfer & Sayatz 2016; Frey & Masiero 2018). Freywald (2016) provided further qualitative research on parataxis in clauses introduced by less studied concessive conjunctions, such as *während* and *wo(hin)gegen*, as well as potential paratactic structures following *dass*. However, research into the phenomenon in digital writing – colloquial language, but not spoken – is still nascent (see Wolfer, Müller-Spitzer, & Ribeiro Silveira 2019), and there have not been any extensive quantitative or longitudinal studies of the phenomenon.

Breindl, Volodina, & Waßner (2014: 841) (in agreement with Volodina (2007)), claimed that verb-second structures appear in written language for only two reasons:

1. In the reproduction of direct speech, or;
2. As part of a wider strategy to give a written text "conceptually oral" features (see Koch & Österreicher 1985), e.g. the use of 'shouting' capitals or repeated letters to represent speech prosody [see Chapter 2, Section 2.1.1.1].

The first point is demonstrated in the only three examples of parataxis identified in the *DWDS-Kernkorpus 21*. All three occurrences of paratactic structures following a subordinating conjunction that it was possible to identify are instances of direct speech, as shown in Example 2.2.

> **Example 2.2** (J. Schweikle (2000), DWDS-KK21 2000/zeit2000_15_5_2244)
> »**Wenn du schreibst Scheiße** über ihn«, sagt Jarek, der Pole.
> (' "**When you write shit** about him", said Jarek, the Pole.')

If the second point is to hold true, we might expect digital writing, e.g. on internet forums, blogs, and other websites, to have a relatively higher rate of parataxis following subordinating conjunctions, as these genres exhibit many other conceptually oral features (Schlobinski 2005: 132; Kleinberger Günther & Spiegel 2006: 112; Androutsopoulos 2007: 87). However, it is not just informal language online that contains non-standard parataxis, as demonstrated in Example 2.3, from an online report on a political website that is part of the *DWDS WebXL Korpus*:

> **Example 2.3** (www.gbw.at (2016), DWDS-WK b71/b71xhfki)
> Wir wollen, dass Kinder bei einer Sommerakademie mitreden und mitgestalten, **weil Kinder verstehen etwas von Abenteuer und Zukunft.**
> ('We want children to participate at a summer academy, **because children understand adventure and future.**')

Bahlo et al. (2019: 147) stated that the phenomenon is also typical of youth language, so we might further expect to see higher rates of parataxis in the *NottDeuYTSch* corpus than in other spoken or online corpora that draw from the language of a more general population. However, there has been very little quantitative comparative study on such syntactic differences between digital youth language and standard language. Accordingly, Section 5.2.3 compares the rates of parataxis following subordinating conjunctions in the online youth language of the *NottDeuYTSch* corpus with those of other corpora.

Within previous research, there has been little consensus on the situations where parataxis can occur. This is best demonstrated by the dispute between Antomo & Steinbach (2010) and Reis (2013) about *weil* clauses, the most examined subordinating conjunction regarding parataxis. Antomo & Steinbach (2010) claimed that the verb-second property of the clause causes semantic, pragmatic and syntactic "disintegration" from the main clause, i.e. the verb-second *weil*-clause then functions as an independent speech act, separate from the main clause (Antomo & Steinbach 2010: 24-25), whereas the use of hypotaxis would result in a "strongly linked" pair of clauses. This is demonstrated to some extent in Example 2.4, which contains both a hypotactic

(in italics) and paratactic (in bold) *weil* clause.

> **Example 2.4** (2014)
> […] ich werde aufjeden fall mir die sachen nal angucken die du gezeigt hast *weil du mein vorbild bist* ich hoffe ich werde bei aowas [sowas] gewinnen **weil du bist echt hüpsch und Nett und du bringst mich immer zum lachen**🙈♡
> ('[…] Ofcourse I will take a look sometime at the things that you showed *because you are my idol* I hope I will win at one of these things **because you are really cute and nice** and you always make me laugh 🙈♡')

The hypotactic clause (italicised) is prototypical for a *weil* clause, demonstrating a strong causal relationship between (potential) effect and the cause. Here the effect is the noun phrase and relative clause *die Sachen […], die du gezeigt hast* ('the things that were shown') and the cause (the reason for following the *YouTuber*'s recommendations) is *weil du mein vorbild bist*, i.e. the fact that the *YouTuber* is the commenter's idol and therefore a trusted influence. In the paratactic *weil* clause, by contrast, the link between effect and cause is less strong, as giving compliments to the *YouTuber* (*du bist echt hübsch […]*), is not a strong causal reason for being selected to win the competition (*bei sowas gewinnen*), although it does justify why the writer might hope to win.

Reis (2013: 225) argued that, rather than the syntactic structure of the clause determining the integration of the subordinate clause and the main clause, it was the "semantics of *weil*" that were important in determining syntax structure. Reis (2013) presented two main hypotheses. Firstly, Reis claimed that if *weil* can be replaced by *denn*, a coordinating conjunction, then parataxis is liable to occur. The following examples from Scheffler (2005: 216) illustrate the differences between *weil* and *denn*, which can both be translated as 'because' in English (also see Pasch et al. 2003: 21-32). *Denn* can be used under the following circumstances:

1. The clause expresses "causation of epistemically judged propositions or speech acts" (Example 2.5).

**Example 2.5** (Scheffler (2005: 215))
Es hat geregnet, **denn** die Straße ist ganz naß
('It has rained, **for** the street is completely wet')

2. The clause is not "a direct answer to a why-question" (Example 2.6).

**Example 2.6** (Scheffler (2005: 216))
Warum ist die Katze gesprungen? — * Denn sie sah eine Maus
('Why did the cat jump? – * Because it saw a mouse')

3. "The content of the because-clause has not been previously mentioned", nor can the *denn*-clause precede the main clause (Example 2.7).

**Example 2.7** (Scheffler (2005: 216))
* Denn es hat geregnet, ist die Straße naß.
* ('Because it rained, the street is wet.')

Secondly, Reis (2013: 227) argued that paratactic syntactic structure can only occur if the 'subordinate' clause has "illocutionary autonomy". The concept of illocutionary autonomy is grounded in the framework of illocutionary acts, developed by Austin (1962) and later expanded by Dore (1975) and Searle (1976), which sought to classify the unspoken implications of a speech act. Searle (1976: 10–16) presents five categories of illocutionary act:

1. Representatives (or assertives): An act committing "the speaker to the truth of the expressed proposition", e.g. stating something.
2. Directives: An act by the speaker to "get the hearer to do something", e.g. asking, ordering, or advising an action.
3. Commissives: An act by the speaker to commit "to some future course of action", e.g. promising to do something.
4. Expressives: An act expressing "the psychological state" of the speaker surrounding an event, e.g. thanking, congratulating, or apologising for something.
5. Declarations: An act that, when successful, "brings about the correspondence between the propositional content and reality", e.g. declaring war, marrying somebody

Searle's approach to classification has since been criticised, primarily for the "assumption that each speech act belongs only to one category [which] fails to account for the multifunctionality of language use" (Nastri, Pena, & Hancock 2006: 1030). However, this does not undermine the analysis proposed by Reis (2013), where a more pragmatic approach to illocutionary acts has been taken. Reis (2013: 224) only identified sentences with "assertive illocutionary potential" (Reis 2013: 224), and the potential that any clause or sentence could have more than one illocutionary aspect was not important, as long as one of the aspects was assertive.

Illocutionary autonomy is therefore found when the supposed subordinate clause contains a self-contained illocutionary act, regardless of or separate from the main clause. As shown in Example 2.8, it could be argued that the paratactic *weil* clause (*weil du bist echt hüpsch und Nett*) is framed as an illocutionary act that contains both directive and expressive aspects. The commenter states that the *YouTuber* has a positive effect on their life (expressive), and through these compliments (as well as the separate hedged directive illocutionary act 'I hope I will win […]'), they wish to be chosen as the winner of the competition.

> **Example 2.8** (2014, repeat)
> […] ich werde aufjeden fall mir die sachen nal [mal] angucken die du gezeigt hast *weil du mein vorbild bist* ich hoffe ich werde bei aowas [sowas] gewinnen **weil du bist echt hüpsch und Nett und du bringst mich immer zum lachen**🙈♡
> ('[…] Ofcourse I will take a look sometime at the things that you showed *because you are my idol* I hope I will win at one of these things **because you are really cute and nice** and you always make me laugh 🙈♡')

The competing hypotheses of Antomo & Steinbach (2010) and Reis (2013) are tested on the data of the *NottDeuYTSch* corpus in Chapter 6, Section 6.2, to determine whether paratactic syntax structure is influenced by lexical choice within the clause, in particular the part of speech of the first field constituent in the paratactic clause.

There has also been significant debate on the grammatical function of the conjunctions that introduce paratactic clauses. Freywald (2016), in a synthesis of the hypotheses of both Antomo & Steinbach (2010) and Reis (2013), argued that *obwohl*, *wobei* und *während*, are syntactically and pragmatically disintegrated as well as illocutionarily autonomous from the main clause. Therefore, they function as coordinating conjunctions, similar to *denn* (Freywald 2016: 31), and should be analysed as such, as there is no subordination occurring. This built on earlier work by Günthner (1996), who claimed that in acquiring the additional function as a coordinating conjunction, these conjunctions had undergone a process of "grammaticalization", i.e. losing some lexical meaning to instead express further grammatical relationships between the content of the clauses. In this case, Günthner (1996) stated that the conjunctions lose some of the strength of the expression of causality (for *weil*) or concession (for *obwohl* and *wobei*) in return for a "discourse-pragmatic meaning", i.e. to organise the discourse or introduce new knowledge instead of supporting the content in the main clause (Günthner 1996: 353-354).

In addition to the grammaticalization of the subordinating conjunctions to coordinating conjunctions, Günthner (1999, 2000a), Gohl & Günthner (1999), and later Frey & Masiero (2018), argued that *weil* and *obwohl* had also undergone a further process of grammaticalization to function as a discourse marker. As opposed to coordinating conjunctions, which allow the combination of two equal, yet syntactically independent clauses to form a "syntactically complex expression" (Leibniz-Institut für Deutsche Sprache 2017a), a discourse marker has four functions, according to Gohl & Günthner (1999: 42) and Blühdorn, Foolen, & Loureda (2017: 24–25), which do not link it syntactically to a previous clause and so it remains syntactically simple. These four functions are to:

1. Introduce additional information
2. Introduce a narrative sequence
3. Introduce a change of topic
4. Structure a conversation

Auer (1997: 55) argued that discourse markers operate semantically separately

from the subsequent clause, although they are syntactically in the same clause, occupying the "Vor-vorfeld" ("pre-first field" position, which would now be described as the left margin, Leibniz-Institut für Deutsche Sprache 2017b). However, Mroczynski (2012) and Frey & Masiero (2018) argued that discourse markers have no fixed syntactic position within the topological model of a clause: they occur "both initially and finally" (Mroczynski 2012: 123), or "outside of the syntactic structure of the clause" (Frey & Masiero 2018: 71). The potential usage of subordinating conjunctions as coordinating conjunctions and discourse markers is examined using the data in the *NottDeuYTSch* corpus as part of the investigation into false positives and ambiguous syntax structures in Section 5.1.3 below, and also analysed in depth in Chapter 6, Section 6.2.1. Digital writing has rarely provided the data for the investigation of these phenomena, as the sources for analysis of discourse markers have overwhelmingly been spoken language. Indeed, both Gohl & Günthner (1999: 47–48) and Mroczynski (2012: 123) observed that discourse markers can be identified in spoken language as they "are often separated from the reference utterance by their own intonation phrase and/or by a pause" (Mroczynski 2012: 123). In written language, discourse markers are less readily identifiable. They may be indicated by punctuation following the conjunction, such as a comma or dash, but for the *NottDeuYTSch* corpus this can also not be relied upon, as the corpus is characterised by a lack of standard punctuation. These challenges are addressed in Chapter 6, Section 6.2, within a wider examination of the grammaticalization of conjunctions in the *NottDeuYTSch* corpus, as one of the potential reasons explaining paratactic clause structures.

### 2.2.5 Orthographic applications

As discussed above in Section 2.1.1, Chapter 7 examines studies on orthographic variation, in particular the use of non-standard spellings to alter the effect of the message (Androutsopoulos 2000), including the representation of conceptually oral features (Dürscheid, Wagner, & Brommer 2010) and discursive strategies (e.g. for politeness, Neuland, Könning, & Wessels 2018). In addition to the research listed above, the analyses in this study also draw

heavily on German language research on youth language specifically in digital contexts, to which I now turn in Section 2.3.

## 2.3 The language of young people in Digitally Mediated Communication

As noted at the start of this chapter, there is an increasing overlap between research on DMC and research on young people, due to their experiences of digital communication technologies from a very early age (Bahlo et al. 2019: 80). In much of the scholarship on the linguistic aspects of DMC (as reviewed in Section 2.1 above), it has been assumed that the language use studied is that of young people, although often this has not been made explicit (as noted in Androutsopoulos 2003b: 189; Siebenhaar 2018: 761), as much data collected online is ethnographically sparse, and researchers have not been able to conclusively establish the demographics of their data sources, often adding a disclaimer (e.g. Arens 2019). The assumption that Digital Writing is produced by young people is partly due to the association of 'new' media usage with young people (Siebenhaar 2006b: 228), and the presence of certain linguistic features supposedly characteristic of youth language, such as the deliberate use of non-standard orthography in written DMC (Kleinberger Günther & Spiegel 2006: 102). However, these features have been attested in the language used by older demographics too (i.e. everyone not counted as "youth," see Siebenhaar 2006a; Kessler 2008). Indeed, Dürscheid (2006: 126) queried whether the non-standard features found in youth language are produced as part of performing youth identity or whether they are simply a characteristic of the online medium, which happens to be particularly frequently used by young people. It is important, therefore, to consider possible differences in the use of similar linguistic features by young people and by a more general population online. This section narrows the focus to research of online language where young people are explicitly stated to have produced the data, as is the case with the data in the *NottDeuYTSch* corpus analysed in this study.

The idea that young people are "immer vernetzt" ('always networked') (Autenrieth & Neumann-Braun 2016: 339) has led to significant scholarship on potential linguistic developments within online youth contexts. Kleinberger Günther & Spiegel (2006: 107–13) suggested that online, young people tend to use and adopt more DMC-specific linguistic features than older demographics, for example, the frequent use of acronyms, phonetic spelling, and the repetition of punctuation or graphemes as laid out by Crystal (2004, 2006) and Jones & Hafner (2012). However, no demonstrable differences have been identified between the orthographic competencies in Digital Writing of adults and those of young people, despite concerns voiced in the media that young people's competence to write standard German is being negatively impacted by digital technologies. Kleinberger Günther & Spiegel (2006), Brommer (2007), Dürscheid, Wagner, & Brommer (2010), Massa (2016), and Wagner & Kleinberger (2016) have all addressed this topic, generally agreeing with the conclusion of Kleinberger Günther & Spiegel (2006: 114) that "Die Rede von einem Verfall der Schreibung sehen wir [...] als Mythos an" ('We [...] regard the talk of a decline in writing as a myth'). There are, however, orthographic differences between Digital Writing of adults and young people, particularly in the frequency of the use of certain features, with young people using emoji more frequently (Mostovaia 2018; Siebenhaar 2018), as well as making more creative use of punctuation (e.g. Busch 2017; Androutsopoulos & Busch 2021). Indeed, these differences have been shown to be conscious decisions that reinforce a young person's identity. Busch (2018) and Androutsopoulos & Busch (2021) presented interview data to show that young people demonstrate a metalinguistic awareness of the distinction between informal and formal writing contexts, particularly regarding the use of standard punctuation, even if they often use non-standard or creative punctuation within their peer groups.

Technological development has been an important spur to new scholarship on the language of young people in DMC. Initial studies used data from text message exchanges (e.g. Höflich 2003; Nowotny 2005), before shifting to popular sites of digital communication used by young people during the 2000s, including instant messaging clients, such as MSN and ICQ, and forums such as

*bravo.de* and *schülerVZ* (e.g. Kleinberger Günther & Spiegel 2006; Dürscheid 2006; Schwencke 2012).[6] From the late 2000s, the rise in smartphone ownership amongst young people and falling data costs in Germany led to a shift in the prominent areas of interaction for young people.[7] This shift has subsequently been reflected in research, with social networking sites, such as *Facebook*, and the mobile messaging client *WhatsApp* becoming sites of considerable interest to academics (for studies on *Facebook*, see Vogelgesang 2010; Brommer & Dürscheid 2012; Dürscheid & Brommer 2013; Androutsopoulos 2015; Voigt 2015; Saxalber & Micheluzzi 2018; Glaznieks & Frey 2018; Vasiljevič 2018; for studies on *WhatsApp*, see Dürscheid & Frick 2014; Busch 2018; Siebenhaar 2018; Koch, Romero, & Stachl 2020; Beißwenger et al. 2020). It is therefore important to further expand the scope of previous research by re-examining it within a new context, including by using additional kinds of data that have not been frequently analysed in existing scholarship, and taking longitudinal approaches to data analysis to examine change over time. Accordingly, the *NottDeuTYSch* corpus was created to allow the examination of authentic youth language produced in *YouTube* comments over a ten-year period from 2008 to 2018 and the re-examination of existing scholarship in the field.

## 2.4 The language of young people on *YouTube*

Having introduced DMC, youth language, and the combination of the two, this section narrows the focus to the language of young people on *YouTube*, the source of the data for the *NottDeuYTSch* corpus, which provides the data for this study. Digital youth language on *YouTube* was only tangentially mentioned in research until 2015. For example, Androutsopoulos (2013: 57), in a study on dialect on *YouTube*, noted the importance of Berlin dialect "to participate efficiently in the capital's youth culture". There is still little research in the sub-

[6] While there was English-language research on the language of young people in chatrooms and on IRC (e.g. Abbott 1998; Sefton-Green 1998; Merchant 2001), German-language research on these platforms tended to focus on dialect usage (e.g. Aschwanden 2001; Siebenhaar 2006c, 2006b) or DMC features (e.g. Rosenau 2001), rather than youth language.

[7] According to Statista (2020), smartphone ownership among 12-19 year-olds increased from 25% in 2011 to around 97% in 2018. Concurrently, the effective cost per gigabyte of mobile data fell from €80 in 2011 to €6.66 in 2018 (Teltarif 2012, 2019).

field of the language of young people on *YouTube*. Studies of the language of *YouTube* comments outside of a German-language context are also rare. In the few studies that exist, the primary focus has been on the performance of intersectional youth identities through orthographic choices (e.g. Mexican-American rap fans, Cutler 2018) and lexical choices as part of metalinguistic discussion (e.g. Arabic-speaking rap fans, Ben Moussa 2019; trans* identities, Ingram 2019; Auxland 2020).

Among the few studies of German-language *YouTube* comments, Meer (2018) examined the performance of youth identity in the comments underneath the videos of a prominent German-language *YouTuber*, Dagi Bee. One of the focuses of Meer (2018: 304–8) was the analysis of parasocial interaction between the commenters and the *YouTuber*.[8] Voigt (2015) conducted a sociolinguistic analysis of *YouTube* videos (alongside *Facebook* and *schülerVZ* posts) created and uploaded by 12-16-year-old German-speaking girls between 2008-2010 to identify linguistic features characteristic of schoolgirls' German, including idiosyncratic spellings of forms of address and "relationship phrases" (e.g. < babiih >, for 'baby'), a high frequency of emoticons, and the use of hedges "to create communicative proximity" (e.g. *eigentlich*, *irgendwie*). However, such studies remain rare. The relative lack of linguistic studies of *YouTube* language further justifies the construction of the *NottDeuYTSch* corpus and the broad scope of the present study to focus on aspects of each of the linguistic fields of lexis, morphology, syntax, and orthography.

## 2.5 Using *YouTube* comments to study digital youth language

This chapter has demonstrated the broad range of topics covered in existing scholarship in the field of DMC, youth language, and digital youth language, as

[8] Parasocial interaction is the illusion of face-to-face contact, i.e. the *YouTuber* addresses the audience directly in the video, and commenters write comments addressed directly to the *YouTuber*, but there is no actual reciprocity and interaction. Parainteraction between young people and *YouTubers* has also been examined in the context of advertising (Böckmann et al. 2019) and political contexts (Dander 2020), although these studies do not primarily undertake linguistic analyses of the comments.

well as several gaps in the current body of knowledge, which the analyses in this study address. As shown in Section 2.1, there has been considerable focus on the relationship between Digital Writing and written and spoken language, which has led to the development and refinement of several frameworks of analysis. However, the section also highlights the limitations of these frameworks, particularly the scope of the frameworks to account for the relatively quick evolution of the linguistic features and communication strategies in Digital Writing. Section 2.2 showed that youth language, like DMC, is also characterised by rapid change in some linguistic features, in this case, lexis. The lexis of German youth language has also been the primary focus in previous scholarship, particularly in how certain constructions can become part of more general colloquial language, but there has been a lack of studies in other aspects of language, such as syntax. The relative paucity of sociolinguistic research is also evident in research at the nexus of DMC and youth language, i.e. digital youth language, as shown in Section 2.3. Furthermore, such studies have not tended to examine digital youth language either quantitatively or longitudinally, instead using either small corpora or data collected from too narrow a timeframe for longitudinal analysis. The section also demonstrated that the few existing sociolinguistic studies of digital youth language remain clustered around two platforms, *Facebook* and *WhatsApp*, rather than *YouTube*. The few studies that have focused on the language of young people on *YouTube*, as examined in Section 2.4, have been broadly limited to ethnographic studies and the analysis of interactive practices, rather than focusing on sociolinguistics, despite the calls for more research in this field (see Androutsopoulos & Tereick 2016: 367).

The construction of the *NottDeuYTSch* corpus of *YouTube* comments not only provides an additional data source for linguistic investigation, but the large size of the corpus at 33 million words enables both quantitative and longitudinal, as well as qualitative approaches to linguistic data. Furthermore, the corpus facilitates the re-evaluation of existing work using different methodologies that can reveal emergent developments in digital youth language. This mixed methods corpus-linguistic approach lays the foundation for the overarching research question of this study, i.e. investigating what we can learn from a corpus-

linguistic and in some cases longitudinal approach to digital youth language, in this case *YouTube* comments. This area of study has long been of interest to me as a researcher, having previously assembled smaller corpora of Digital Writing by young people from forums posts on BRAVO.de (Cotgrove 2013) and the comments sections of music videos on YouTube, which served as the pilot study for this thesis (Cotgrove 2017).

While the focus of Cotgrove (2013) was on lexical items, more specifically, Anglicisms, the current thesis seeks to take a broader approach to Digital Writing by young people, covering lexis, morphology, syntax, and orthography. To do so, the thesis draws on the *NottDeuYTSch* corpus to present three case studies that show the value of such an approach and help address specific gaps in existing scholarship in these areas.

In Chapter 4, I use *geil* to build a case study of lexis and morphology. *Geil* has been chosen because it is an often-referenced lexical item within scholarship and in public discourse, but has not yet been examined using the in-depth longitudinal and quantitative approaches enabled by the large size of the *NottDeuYTSch* corpus. The analysis takes a two-pronged approach: examining the change in use of *geil* over time, and the morphological characteristics of different senses of *geil*.

The syntax case study in Chapters 5 and 6 addresses the lack of comparative grammatical studies of online youth language to date by examining the use of non-standard word order following subordinating conjunctions in the *NottDeuYTSch* corpus compared to three other corpora. Taking advantage of the size of the *NottDeuYTSch* corpus, the chapter uses the new data to re-examine previously established hypotheses about which subordinating conjunctions can be used to introduce paratactic clauses, and why.

Finally, the case study of orthography in Part III addresses two gaps in existing scholarship on Digital Writing, specifically concerning metacommunication: firstly, the lack of longitudinal overview of orthographic features that have been used since the 1980s for metacommunication; and the lack of a comprehensive framework to analyse all orthographic features used for metacommunication

in Digital Writing.

Overall, the analysis of research to date on DMC and youth language in this chapter has demonstrated that despite the wide of range of studies in these fields over the past 40 years, there are still some significant gaps in the research. These gaps will be addressed in Chapters 4 to 7 of this study through the linguistic analysis of the *NottDeuYTSch* corpus, which also enables the re-examination and re-evaluation of previously researched topics using new data. The analysis chapters not only identify emergent features outside of the field of lexis, in both morphology and syntax, but also provide previously absent evidence of longitudinal developments of lexical, morphological, orthographic, and syntactic features. In doing so, the study significantly contributes to the research areas of both DMC and youth language, as well as the linguistically important yet under-researched field of digital youth language.

# 3

# Constructing the *NottDeuYTSch* corpus

This chapter presents the process of collecting *YouTube* comments to create the ***Nott****inghamer Korpus* ***deu****tscher* ***Y****ou****T****ube-****S****pra****c****he* ('The Nottingham German *YouTube* Language Corpus' – or *NottDeuYTSch* corpus for short) compiled for the purposes of this thesis to analyse the online language produced by young German-speakers. The *NottDeuYTSch* corpus contains over 33 million words taken from approximately 3 million *YouTube* comments from videos targeted at a young, German-speaking demographic. The *YouTube* comments, which represent an authentic language snapshot of young German speakers, are used to investigate the linguistic features that characterise online German *Jugendsprache*. The corpus was designed in three stages:

1. 112 popular German-speaking *YouTube* channels were selected to represent the mainstream *YouTube* scene in the DACH region, using a variety of media sources, as described in Section 3.1.3.
2. Every video uploaded by the 112 channels was collated into a database, eventually consisting of 99,334 videos, referred to as the 'pre-corpus database'. The metadata of these videos included the time of upload,

number of comments and video category. The time of upload provides a chronological overview of the distribution of videos; the number of comments indicates engagement from the target population; and the video category provides a broad content description, selected by the uploader from a list of 31 options, such as Gaming, Entertainment, or Sports.[1]

3. Extracting every comment from the 99,334 videos was not feasible, so videos were selected to provide the comments for the *NottDeuYTSch* corpus, proportionally sampled from the pre-corpus database based on the number of comments written per video category and per year in the period 2008-2018.

Constructing the corpus in this way ensured that genre and time-based differences are taken into account when examining the corpus as a whole, and also genre and longitudinal analyses, i.e. comparing language use across video categories and language change over the time period. The corpus also contains a considerable amount of associated metadata for each comment that enables further longitudinal analyses.

This chapter is divided into four sections. Section 3.1 presents the methodology behind selecting the data for the *NottDeuYTSch* corpus, including the guiding principles of building the corpus and identifying the *YouTube* channels from which the comments were collected. Section 3.2 outlines the processes of constructing the *NottDeuYTSch* corpus, examining methodological concerns, such as corpus balance and size, and gives an in-depth explanation of sampling procedures. Section 3.3 provides an overview of the *NottDeuYTSch* corpus and contains a breakdown of the key statistical features and finally, Section 3.4 outlines the potential applications of the corpus within future linguistic research.

[1]For Germany and Austria, the complete list of 31 categories (translated into English) are as follows: Film & Animation, Autos & Vehicles, Music, Pets & Animals, Sports, Short Movies, Travel & Events, Gaming, Videoblogging, People & Blogs, Comedy, Entertainment, News & Politics, Howto & Style, Education, Science & Technology, Nonprofits & Activism, Movies, Anime/Animation, Action/Adventure, Classics, Comedy, Documentary, Drama, Family, Foreign, Horror, Sci-Fi/Fantasy, Thriller, Shorts, Shows, and Trailers.

## 3.1 Selecting the data

*YouTube* is a significant source of authentic linguistic data created by young people. However, as shown in Chapters 1 and 2, there are significant gaps in corpus linguistic scholarship within the field. This section presents the methodological processes and principles behind selecting the data for the *NottDeuYTSch* corpus. Section 3.1.1 presents the aims and objectives of the corpus, and the typical content of the videos selected to provide comments for the corpus. Section 3.1.2 presents the case for treating the comments collected to construct the corpus can be considered as authentically produced by young people, and the ethical considerations surrounding the data. Finally, Section 3.1.3 details the processes to identify the channels and videos to be included in the pre-corpus database, in preparation for sampling to create the *NottDeuYTSch* corpus.

### 3.1.1 Principles of building the *NottDeuYTSch* corpus

To investigate the language used by young German speakers, videos produced by *YouTube* channels popular with the 12-19-year-old demographic were selected for inclusion in the corpus. Three main factors governed the text selection for the corpus:

1. The *NottDeuYTSch* corpus should represent, as best as possible, the language used by young German-speakers online. As the corpus is only a sample of the German-language *YouTube* content that has been produced, it is impossible to achieve perfect representativeness, although every effort has been made to ensure that the data were selected according to a strict methodology, as presented in Section 3.1. Inevitably the selection process has been guided – and possibly limited – to some extent by my previous exposure to German-language *YouTube* channels and German youth culture.
2. The data must be able to be analysed longitudinally. A preliminary investigation undertaken in preparation for this study suggested that there is demonstrable language change between the online language used by young German-speakers of 2008 and 2018. Therefore, all data down-

loaded to form the corpus are timestamped so that it is possible to make comparisons over the timeframe of the corpus (roughly a decade).

3. The *NottDeuYTSch* corpus must be able to be used in comparison with other German-language corpora. To substantiate what I claim to represent online youth language, I compare language usage statistics with existing corpora that are representative of the general language throughout this study, particularly in Parts I and II, such as the German Reference Corpus (*DeReKo*), corpora that represent general online language, such as the *DWDS Webkorpus*, and corpora containing online youth language, such as the *Dortmunder Chat-Korpus*.

### 3.1.2 The identity of the commenters

The *NottDeuYTSch* corpus is intended to be a collection of authentic language created by German-speaking young people. However, verifying the age of the commenters presents a methodological challenge for the construction of the corpus, as this knowledge is not publicly available and is often not disclosed within a comment. Sometimes a commenter will provide such information, e.g. in Example 3.1, where the commenter declares both an age and gender identity, but this is rare. This problem is addressed through an inferential approach based on the target demographic of the selected channels, reinforced by a post-hoc statistical analysis of explicit disclosures of age within the comments of the *NottDeuYTSch* corpus.

> **Example 3.1** (2012)
> Ich bin kein Boy, net 6.Ich bin 12 und ein Mädchen, also kann ichs ja sagen: Machs doch besser!! Tja, jeder hat halz den anderen Geschmack -.- Ich mag sie <3 ('I am not a boy, not 6. I am 12 and a girl, so I can say this: Do it better!! Well, everyone has different tastes -.- I like her <3')

Directly asking every single commenter their age to verify that the comments are mostly written by young people was of course neither feasible nor appropriate. Instead, the thesis follows approaches suggested by Döring (2010) that describe how an online user may present their digital identity to infer that the

comments are generally written by young people. Döring (2010: 164) suggests that a person presents their identity online in five different ways:

1. Direct self-presentation: Information that explicitly presents their identity, such as photos and statements.
2. Indirect self-presentation: Information that implicitly presents their identity, such as linguistic cues (e.g. dialect, technical terms, spelling errors).
3. Third-party personal presentation: Information, such as photos and statements, about a person disclosed through a third party.
4. Personal presentation through media: Information, such as interviews and other digital traces, that disclose a person's identity.
5. Personal presentation through IT-systems: Information (and meta-information) on the user created as a result of their online activity, such as login and sign-up dates to various platforms.

A combination of direct self-presentation and personal presentation through IT-systems was used to establish the rough age bracket of the commenters in the *NottDeuYTSch* corpus.[2] Direct self-presentation is when a commenter explicitly states their age in a comment, as in Example 3.1 above. The corpus was searched for statements, such as "ich bin XX" or "ich bin um die XX", where *XX* is a two-digit number. There were 7,121 comments (0.23% of all comments) identified that directly stated the age or age group (e.g. child or adult) of the commenter. Of these comments, 94% (6,729) provided an age under 19 or a birth year of 2000 or later, which would guarantee an age of under 19 within the timeframe of the corpus. This would strongly suggest that the vast majority of commenters are young people, even if young people are more likely to disclose their age, reminiscent of Adrian Mole (aged 13 $\frac{3}{4}$).

The other approach suggested by Döring (2010) – personal presentation through IT-systems – used to infer the age bracket of the users by examining the choice of video commented on. The videos were specifically selected to contribute to the corpus because they were produced to target the 12-19-

[2] Indirect self-presentation through the analysis of language could not be used to identify young people in this thesis, as linguistic features of young people are the subject of the investigation, which would create a recursive, self-referential, circular loop.

year-olds by *YouTube* channels that consistently and successfully target this demographic (as explained fully in Section 3.1.3 below). Therefore, the corpus reflects German-language youth culture. While there may be commenters who would not be counted as young people, the small size of this group, roughly 5% based on the self-disclosure statistics, would not significantly statistically affect the analyses. The audience, and therefore, those commenting, can be overwhelmingly considered as part of the demographic in question, and the corpus is thus suitable for research on the language of young people.

Having established that the comments are highly likely to have been written by young people, there are some ethical considerations regarding safeguarding that it was necessary to address. Following the advice of Marx & Weidacher (2020) on digital privacy, since the commenters are likely under the age of majority, all efforts have been made to anonymise the commenters and remove all identifying information, such as place names, phone numbers, and addresses, despite the comments being publicly accessible.

### 3.1.3 Identifying relevant *YouTube* channels

The following section presents a detailed examination of the process to identify and select the *YouTube* channels that comprise the pre-corpus database, from which the comments were sampled to build the *NottDeuYTSch* corpus. The content creators behind *YouTube* channels have diverse media backgrounds: aside from channels belonging to traditional media companies, there are amateur vloggers (video bloggers), amateur-turned-professional vloggers, who earn money through advertising revenue and sponsorships, and other agents that use *YouTube* to advertise themselves or their brand, such as musicians and corporations. Diversity among content creators has existed on *YouTube* since its early days (Burgess & Green 2018: vi-vii), but a large proportion of popular channels now belong to professional *YouTube* vloggers, who produce videos that fall under the new industry sector of "social media entertainment" (SME). The term SME refers to a broad range of content that can be defined as "vlogging, gameplay and do-it-yourself (DIY) style and beauty" (Cunningham & Craig

2017: 71-72). The phenomenon is in part due to two commercial structural changes to *YouTube*. Firstly, *YouTube* opened access to the *YouTube Partner Programme* to all users in 2012, which enabled channels over 100 subscribers to claim a share of the advertising revenue on adverts shown when people watched their videos (Burgess & Green 2018: 55). Secondly, the creation of the paid subscription service, *YouTube Red*, in 2015 (rebranded as *YouTube Premium* in 2018) offers advert-free streaming and access to exclusive content from popular channels and *YouTubers*, such as 'Scare PewDiePie' with PewDiePie (110 million subscribers as of 2021) or 'Foursome' with Jenn McAllister (3.5m subscribers) and Logan Paul (23.2m subscribers).

The popularity of SME is widely appreciated among young people who watch *YouTube*, with eight of the ten most popular German-language channels producing SME content.[3] These channels from professional *YouTube* stars often have the highest number of subscribers, views, and videos on *YouTube*, and the videos they upload attract many comments, which is reflected in the large proportion of the comments under their videos that are included in the *NottDeuYTSch* corpus. The stars of the channels also enjoy offline success, appearing in other teen-orientated media, such as the long-running German-language teen magazine, *BRAVO*. The magazine has been an important outlet for youth interests since its first publication in 1956, focusing on film, music, and celebrities, as well as coming-of-age topics such as relationships and sexuality. Since August 2014, German-language *YouTubers* have been regularly featured on the front cover of the print magazine (see Figure 3.1, left). At the end of 2014, *BRAVO* consolidated its interest in *YouTube* by launching the 'YouTube' section of their website (https://www.BRAVO.de/youtube) and a quarterly magazine called '*BRAVO Tubestars*' (see Figure 3.1, right), which featured interviews with German-language *YouTubers* and reports on their videos. As explained below, a large number of *YouTube* channels selected to construct the corpus featured in *BRAVO* publications.

[3]The 10 German-language channels on *YouTube* with the most subscribers, as of February 2019, were FreeKickerz (Sports, 7.5m subscribers), BibisBeautyPalace (SME, 5.6m), Julien Bam (SME, 5.2m), Gronkh (SME, 4.9m), The Voice Kids (Music, 4.4m), Simon Desue (SME, 4.2m), Dagi Bee (SME, 4.0m), ApeCrime (SME, 3.8m), Julienco (SME, 3.8m), Kinder Spielzeug Kanal (SME, 3.8m).

Figure 3.1: Left: BRAVO magazine cover, 29.08.2014, with Bianca Heinicke (Bibi) (top centre), Y-titty (bottom right), and ApeCrime (bottom left). Right: 'BRAVO Tubestars' cover, 29.10.2014, featuring popular *YouTubers*

The *NottDeuYTSch* corpus is not solely comprised of SME-orientated channels; the aim in compiling the corpus was to include channels with as wide a range as possible of popular youth-focused interests to represent what young German-speakers are watching and commenting on. A large proportion of videos on *YouTube* belong to existing media institutions, such as radio, television, and print companies. These companies use *YouTube* to advertise their other media platforms by syndicating content on *YouTube*. For example, *1LIVE*, the youth radio station of *WDR*, uploads their short animation series, *Babo-Bus*, to both *YouTube* (1Live 2019a) and the 1LIVE website (1Live 2019b) but includes links in the *YouTube* video description to other *1LIVE*-owned websites and social media platforms (see Figure 3.2) to generate further interest. The content produced by these media institutions is mostly German-speaking, which is reflected in the high use of German by the commenters.

Selecting (or rejecting) channels and their videos based on the language of the comments was a methodological consideration that required careful deliberation. The corpus is designed to examine the language of young German-speakers, so only videos where German is the primary language of the video were selected. English-language videos, although sometimes aimed at a German audience, tend to attract a worldwide audience that speak varieties of English. While many young German-speakers may have the English-language proficiency to interact with English-language internet culture, investigating how young people communicate in online environments targeted at German speakers (and the linguistic variation within them) is the focus of the research presented in this thesis, rather than a general overview of how German-speakers communicate online. This fine line is reflected in the relatively low amount of music videos that met the selection criteria for the corpus, despite the extremely large numbers of German-language musicians and music industry institutions that upload content to *YouTube* and the large number of views and comments that these videos garner from young people.[4] A large number of music videos popular amongst young German-speakers are sung in English and contain a

[4]According to the Music Consumer Insight Report published by the International Federation of the Phonological Industry in 2017, 46% of all music streamed was via *YouTube* (IFPI 2017: 5) and 98% of internet-using 16-24 year-olds consumed music online (IFPI 2017: 7).

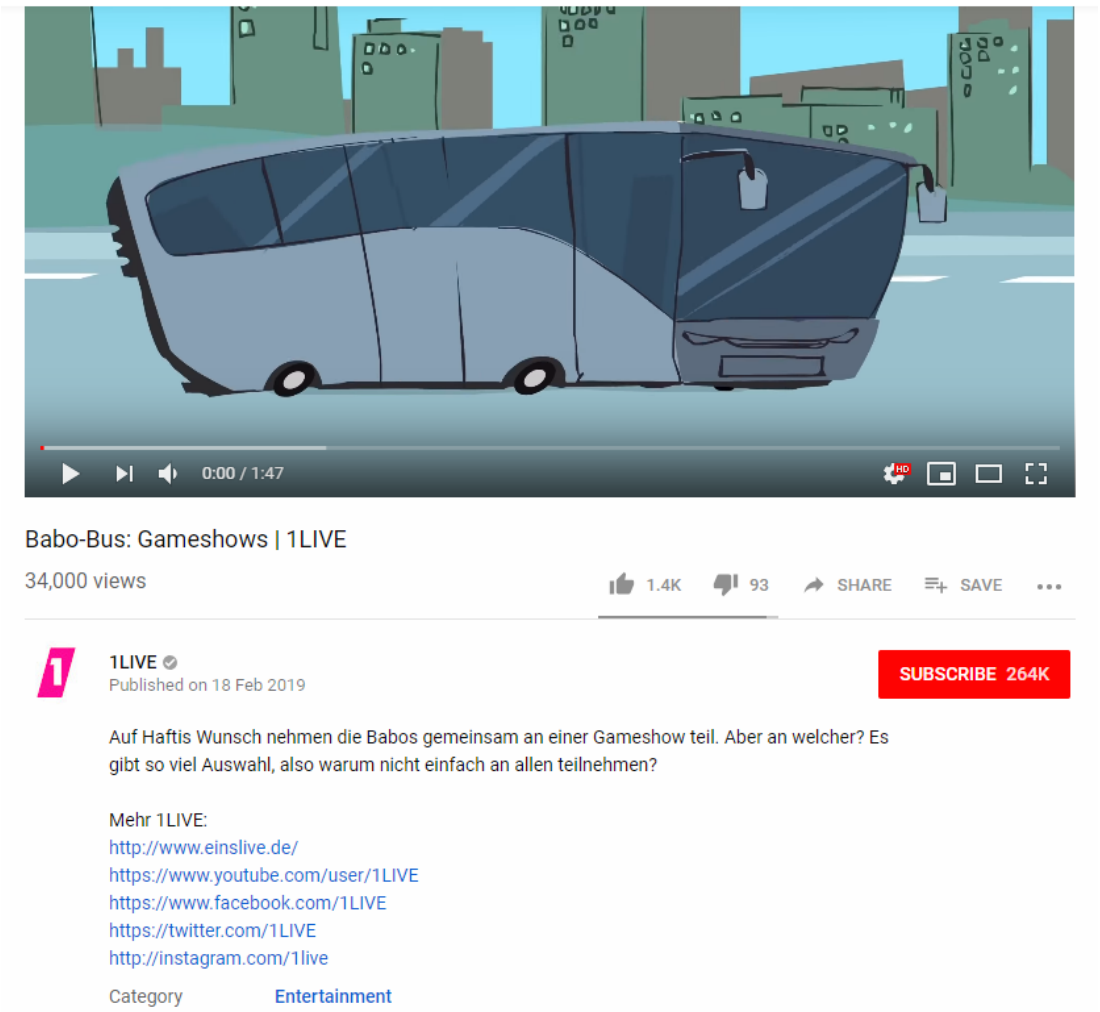



Figure 3.2: 1LIVE Babo-Bus: *YouTube* video description contains links to other online platforms belonging to the uploader

large proportion of English-language comments, and so were not selected for inclusion in the corpus. This conscious choice to only select videos that use German as the primary language also excludes videos with a large amount of German-language comments from content creators with a German-speaking background who upload in English. For example, the *YouTube* channel from Hamburg-born singer Sarah Connor is excluded from the corpus because of the large proportion of English-language songs uploaded to her channel and the subsequent high volume of English-language comments underneath the videos, despite her popularity in German-speaking countries.

The pre-corpus database contains 112 *YouTube* channels that cover the most popular topics and videos watched by young German-speakers. While there were more than 112 *YouTube* channels that produce relevant content for the

*NottDeuYTSch* corpus and have a young, German-speaking audience, they cannot yet be described as mainstream, as they are smaller channels and do not yet attract a sufficient number of subscribers, views, or most importantly, comments for inclusion in the corpus. As a rough guide, channels that had at some point surpassed 100,000 subscribers were considered eligible for inclusion in the corpus, as a channel with this number of followers qualifies for the Silver Creator Award from *YouTube*, a benchmark for mainstream channel success (YouTube 2021).

The process of identifying the channels was informed by my previous exposure to German-language *YouTube* culture. The channels identified had either received considerable media attention due to their *YouTube* popularity, such as *BibisBeautyPalace*, or were owned by media companies specifically targeted at young people, such as the *YouTube* channel of the radio station *1Live* (the youth station of *WDR*). I started collecting background information on Bibi from *BibisBeautyPalace* and noticed that she often appears on the front cover of *BRAVO*. Due to the magazine's prominent role in German-speaking youth culture and regular news items involving German-language *YouTubers*, 63 of the 112 *YouTube* channels in the database featured in *BRAVO* cover stories and website home page articles.

While *BRAVO* is an important source of youth culture, it would not have been methodologically sound to restrict my database to *YouTube* channels featured online or in print media published by *BRAVO*. *BRAVO* focuses solely on featuring individual *YouTubers*, rather than *YouTube* channels run by youth media outlets or music artists, which often upload videos that garner far more views than the average video of a *YouTuber* and a similar average number of comments.[5] To address this, I added music channels to my database by analysing the German music charts for successful German-speaking artists over the past 10 years and German music *YouTube* channels, such as *AggroTV*, the online presence of the former hip-hop record label, *Aggro Berlin*. *AggroTV* has been broadcasting online since 2006, has accumulated over 1.1bn views among its 3,500 videos, and

[5] Based on my database, the average number of views per video for music channels is roughly three times higher than for videos uploaded to a typical individual *YouTuber*'s channel.

has collaborated with almost every hip-hop star of note in Germany and Austria at some stage since its inception. Eight artist or music channels were selected who had the highest chart success, largest *YouTube* presence, and highest number of appearances in youth media (including *BRAVO*). The music channels selected have, on average, 380m views per channel, and around 2,000 comments per video.

In addition to the music channels, I selected five successful youth/online media platforms that have a high number of views and subscriptions on *YouTube*, such as PromiFlash, the leading *YouTube*-based news service aimed at young German speakers. Many traditional broadcasters maintain an online presence, such as *Der Spiegel* (https://www.youtube.com/user/spiegeltv), SWR (https://www.youtube.com/user/SWR), or *Bayerischer Rundfunk* (https://www.youtube.com/user/BR), but they enjoy varying levels of success and often do not engage with young people, instead focusing on specialist topics that accompany publications or programmes broadcast through their main medium. Each of the five media outlet channels included in the corpus (*1Live*, *AGGRO.TV*, *Promiflash*, *RTL*, *World Wide Wohnzimmer*) has at least 50m views and has uploaded 500 videos, with *PromiFlash* leading the way with almost 2bn views and 1.2m comments.

To verify that my existing database was representative of the most popular *YouTubers* in the German-speaking area, I used a *YouTube* social aggregation website *SocialBlade* which lists the 250 channels in each of Germany, Austria, and Switzerland with the most subscribers. This would verify whether the channels featured in *BRAVO* had a large enough number of subscribers to be eligible for inclusion in the corpus. Using the Internet Archive (https://archive.org/web), I was able to observe how this list had changed since 2014, which helped identify *YouTube* channels aimed at the demographic that were popular in the past. This was crucial to ensuring that the *NottDeuYTSch* corpus is as representative as possible of all years encompassed by the corpus, not just at time of its construction. This process added nine further channels to the database, such as Coldmirror's channel (https://www.youtube.com/user/coldmirror). Coldmirror's parody videos of Harry Potter and internet culture made her the lead-

ing German *YouTuber* and winner of the young talent award at the inaugural *Deutscher Webvideopreis* in 2011.

The final process in expanding the database was to explore the 'Related Channels' section on the 'About' page (as in Figure 3.3) from the 101 channels in the database identified up to this point. To do so, I used the 'YouTube Tools Channel Network Module' (https://tools.digitalmethods.net/netvizz/youtube/mod_channels_net.php), which produces a list of channels that are similar to, or recommended by, the list of channels inputted. Combined with manual checks of the 'Related Channels' sections to ensure that the programme was working correctly, I added eleven more channels to the database, mostly consisting of smaller, yet still mainstream channels, such as that created by Tina Neumann, from Graz, who first found online fame in 2017 as a 15-year-old on Musical.ly, a site where users lip-sync to songs, before building a sizeable *YouTube* and Instagram following. As the time of the construction of the corpus in Autumn 2018, her channel had 250,000 subscribers and over 18m views on *YouTube*. A breakdown of the sources of the channels for the *YouTube* corpus is presented in Table 3.1, and a list of the selected channels in Table 3.2.

Table 3.1: Breakdown of sources used to identify channels to include in the *NottDeuYTSch* Corpus

| **Channel Identification Process** | **Number of Channels Identified** |
|---|---|
| Existing Knowledge | 7 |
| BRAVO Magazine Covers and Website | 63 |
| Music Channels | 8 |
| Youth Media Channels | 5 |
| SocialBlade.com | 18 |
| Related Channels | 7 |
| YouTube Tools Channel Network Module | 4 |
| Total | 112 |

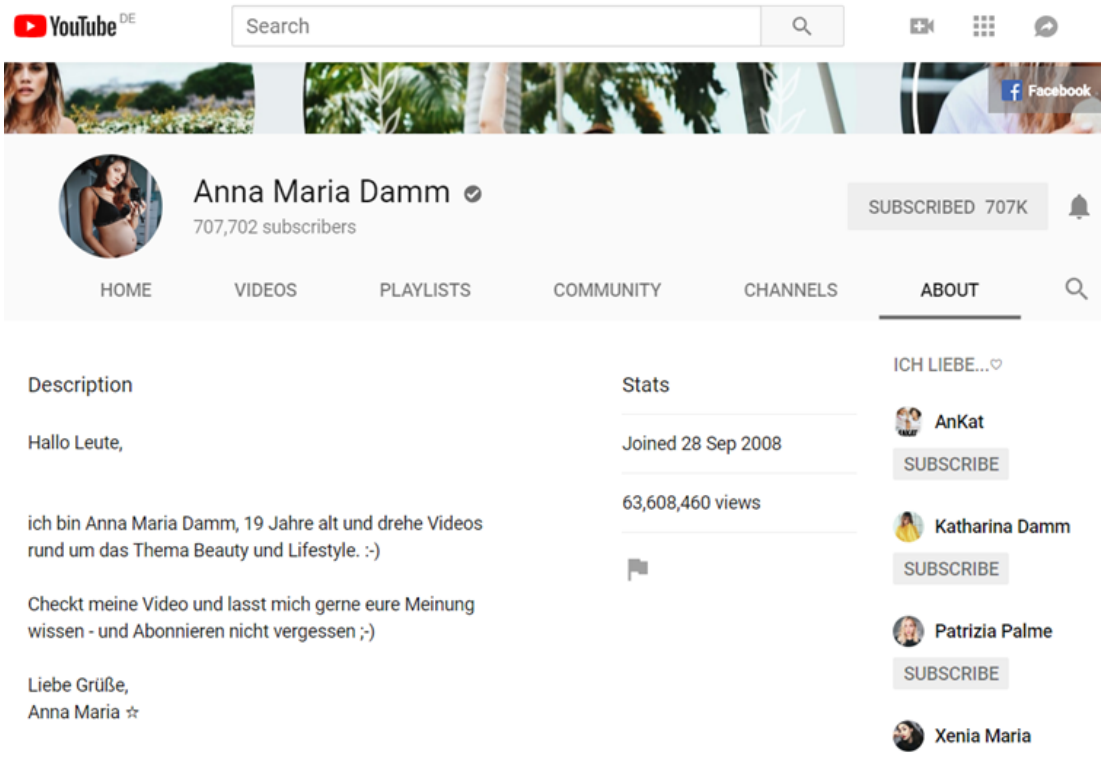



Figure 3.3: *YouTube* channel 'About' page, showing channel description, number of subscribers, the date that the channel was created, related channels (curated by the channel owner), and the total number of views of all the videos in the channel

## 3.2 Constructing the *NottDeuYTSch* corpus

This section explains the methods taken to construct the *NottDeuYTSch* corpus. Section 3.2.1 outlines the process of extracting and cleaning of the data. Section 3.2.2 outlines the steps taken to ensure the corpus is as balanced and representative as can be. Section 3.2.3 explains how the corpus can be considered an appropriate size for the linguistic analyses undertaken in Chapters 4 to 7.

### 3.2.1 Extracting and cleaning the data

The pre-corpus database was imported into the statistical software, R (R Core Team 2021), for pre-processing. This consisted of writing code to interact with the *YouTube* Application Programming Interface (API), to import data on the channels and videos effectively (instead of manually copying and pasting data).

Table 3.2: List of channels included in the *NottDeuYTSch* corpus

| **YouTube channel names** | | | |
|---|---|---|---|
| 1LIVE | DieLochis | Lamiya Slimani | Sami Slimani |
| AGGRO.TV | Dilara Duman | LauraJoelle | Sarazar |
| AlexiBexi | Dima | LeFloid | Shirin David |
| Ana Lisa | E IL o T IR i X ™ | Lena | Simon Desue |
| Anna Maria Damm | Ebru Ergüner | Leon Machère | Sonny Loops |
| Anne Wünsche | Emrah | Lisa-Marie Schiffner | SophiaThiel |
| Anni The Duck | FabTheGap | MRS BELLA | TZON4life |
| ApeCrime | Felix von der Laden | Maren Wolf | TheBeauty2go |
| ApeCrimeTV | FloWest | MefYou | Tina Neumann |
| ApoRed | Freshtorge | Meggyxoxo | Tobias Wolf |
| Applewar | Hatice Schmidt | Melina Sophie | Tugay |
| AshMadeOurEyes | HeyMoritz | Michael Buchinger | ViktoriaSarina |
| BULLSHIT TV | Isabeau | Michael Schulte | World Wide Wohnzimmer |
| BUSHIDO | Ischtar Isik | Mirco Rosik | YTITTY |
| BarbaraSofie | JONAS | MixX - Die Klassiker | albertoson |
| BibisBeautyPalace | Jarow | Nihan | atzenmusiktv |
| Bonnytrash | Joyce | Nilam | bademeisterTV |
| Chameen | Julia Beautx | Niloofar Irani | coldmirror |
| Chimperator Channel | Julien Bam | Paluten | diejungs |
| ConCrafter \| LUCA | Julienco | Pamela Rf | flyinguwe |
| DASDING | KAYEFTV | Paola Maria | iBlali |
| DIE TOTEN HOSEN | KWiNK | PietSmiet | inscope21 |
| Dagi Bee | Katharina Damm | Pietro Lombardi | kitthey |
| DasMirkoo | KeysJore | Promiflash | madametamtam |
| Dennis Bro | KrappiWhatelse | RTL | manniac |
| Der Ömsen | KsFreakWhatElse | RayFox | ungespielt |
| Diana zur Löwen | LIONTTV | Rebekah Wing | xKarenina |
| DieAussenseiter | LIONTVLOG | Sam Masghati | xLaeta |

Following this, the total number of videos and comments could be established for sampling, as well as their distribution across video category and year. The initial size of the pre-corpus database was 102,115 videos, but approximately 3,000 videos were removed as they did not have any comments that could be

extracted, because the uploader either had disabled comments for that video or had streamed the video live through *YouTube*. This brought the total number of videos to 99,334. Whilst comments under a live-streamed video can be extracted using other methods, I chose not to include them as the interaction between commenters and the nature of their participation in a 'live' environment creates a different communicative environment: for example, comments simply express that a user is virtually present, rather than interacting with the content of the video or other users (Stenson 2020: 233).[6]

### 3.2.2 Corpus representativeness and balance

The *NottDeuYTSch* corpus consists of comments that can be analysed and the results generalised to represent the language used by young German-speakers in comments under mainstream *YouTube* videos. If the results from investigating a corpus can be generalised to the total population, the corpus is said to be 'representative' of the population in question (Biber 1993: 243), which in this case is the pre-corpus database. The database contains information on the upload date and video category for every video uploaded by the 112 channels. These two parameters, date and category, were used to ensure the representativeness of the *NottDeuYTSch* corpus: the proportions of videos uploaded per year and per category were mirrored in the eventual selection of the videos for the corpus. Selecting the videos according to video category is an example of stratified random sampling, which McEnery, Xiao, & Tono (2006: 20–21) suggest is the best way of ensuring corpus 'balance' where a corpus contains "a wide range of text categories" (McEnery, Xiao, & Tono 2006: 16). A balanced corpus is important to ensure that (in this case) German online youth language can be appropriately analysed, for example, to account for language variation.

As can be seen in Table 3.3, the median number of comments per video varies considerably between categories, with the 'Autos & Vehicles' category having the most, at over 1,000 comments per video. However, there is no correlation between the median number of comments per video and the number of videos

[6] For more on audience communication in a live-stream environment, see Hamilton, Garretson, & Kerne (2014: 1322–24).

in a category. A further breakdown of the distribution of the videos in the pre-corpus database according to video category and year is shown in Table 3.4. Gaming and Entertainment are the categories with the highest total number of videos and views, but videos in these categories do not consistently generate high numbers of comments or views; instead figures tend to towards the extremes. For example, of the 76,000 videos in the Gaming and Entertainment categories, over 52,000 videos had fewer than 500 comments and 52 videos had over 50,000 comments. Therefore, both the number of videos and number of comments needed to be taken into account when sampling the pre-corpus database.

Table 3.3: Number of videos and associated statistics for each *YouTube* category from the *YouTube* channels in the database

| **Category** | **View Count** | **Comment Count** | **Video Count** | **Median Views per Video** | **Median Comments per Video** |
|---|---|---|---|---|---|
| Autos & Vehicles | 51 993 594 | 242 557 | 113 | 359 728 | 1 065 |
| Comedy | 1 507 207 055 | 4 666 523 | 2 416 | 178 328 | 575 |
| Education | 278 653 164 | 1 504 014 | 248 | 326 739 | 557 |
| Entertainment | 16 277 739 328 | 55 938 493 | 36 354 | 130 834 | 192 |
| Film & Animation | 508 668 833 | 1 294 815 | 848 | 149 038 | 562 |
| Gaming | 6 352 437 868 | 22 638 805 | 39 772 | 82 313 | 236 |
| Howto & Style | 4 886 188 157 | 44 804 498 | 9 281 | 199 563 | 622 |
| Music | 826 540 559 | 2 329 750 | 1 455 | 44 984 | 44 |
| News & Politics | 79 325 301 | 336 344 | 129 | 138 527 | 484 |
| Nonprofits & Activism | 14 643 032 | 73 297 | 65 | 127 718 | 767 |
| People & Blogs | 2 732 357 201 | 11 900 926 | 5 651 | 179 251 | 654 |
| Pets & Animals | 4 798 619 | 31 703 | 19 | 198 184 | 494 |
| Science & Technology | 17 452 464 | 93 092 | 116 | 23 637 | 16 |
| Shows | 1 215 471 180 | 3 078 879 | 1 254 | 260 141 | 542 |
| Sports | 268 256 970 | 898 499 | 985 | 149 468 | 328 |
| Travel & Events | 284 939 433 | 1 056 998 | 628 | 338 312 | 994 |
| **Total** | **35 306 672 758** | **150 889 193** | **99 334** | **113 183** | **272** |

Table 3.4: Total number of videos per category and year in the pre-corpus database

| Category/Year | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Autos & Vehicles | 0 (0%) | 1 (0%) | 3 (0%) | 2 (0%) | 4 (0%) | 0 (0%) | 2 (0%) | 4 (0%) | 29 (0.03%) | 47 (0.05%) | 21 (0.02%) | 113 (0.11%) |
| Comedy | 13 (0.01%) | 99 (0.1%) | 214 (0.22%) | 254 (0.26%) | 270 (0.27%) | 234 (0.24%) | 301 (0.3%) | 263 (0.26%) | 334 (0.34%) | 312 (0.31%) | 122 (0.12%) | 2 416 (2.43%) |
| Education | 2 (0%) | 7 (0.01%) | 23 (0.02%) | 28 (0.03%) | 7 (0.01%) | 4 (0%) | 69 (0.07%) | 57 (0.06%) | 51 (0.05%) | 0 (0%) | 0 (0%) | 248 (0.25%) |
| Entertainment | 127 (0.13%) | 248 (0.25%) | 950 (0.96%) | 1 513 (1.52%) | 1 900 (1.91%) | 1 860 (1.87%) | 2 605 (2.62%) | 3 138 (3.16%) | 5 507 (5.54%) | 11 675 (11.75%) | 6 831 (6.88%) | 36 354 (36.6%) |
| Film & Animation | 3 (0%) | 13 (0.01%) | 60 (0.06%) | 95 (0.1%) | 123 (0.12%) | 96 (0.1%) | 109 (0.11%) | 94 (0.09%) | 95 (0.1%) | 129 (0.13%) | 31 (0.03%) | 848 (0.85%) |
| Gaming | 1 (0%) | 6 (0.01%) | 60 (0.06%) | 1 757 (1.77%) | 5 887 (5.93%) | 9 014 (9.07%) | 7 790 (7.84%) | 6 283 (6.33%) | 4 305 (4.33%) | 2 614 (2.63%) | 2055 (2.07%) | 39 772 (40.04%) |
| Howto & Style | 59 (0.06%) | 301 (0.3%) | 237 (0.24%) | 277 (0.28%) | 647 (0.65%) | 1 178 (1.19%) | 1 419 (1.43%) | 1 363 (1.37%) | 1 458 (1.47%) | 1 614 (1.62%) | 728 (0.73%) | 9 281 (9.34%) |
| Music | 150 (0.15%) | 52 (0.05%) | 141 (0.14%) | 151 (0.15%) | 233 (0.23%) | 83 (0.08%) | 70 (0.07%) | 125 (0.13%) | 94 (0.09%) | 248 (0.25%) | 108 (0.11%) | 1 455 (1.46%) |
| News & Politics | 0 (0%) | 7 (0.01%) | 30 (0.03%) | 28 (0.03%) | 13 (0.01%) | 9 (0.01%) | 7 (0.01%) | 12 (0.01%) | 11 (0.01%) | 7 (0.01%) | 5 (0.01%) | 129 (0.13%) |
| Nonprofits & Activism | 0 (0%) | 3 (0%) | 3 (0%) | 3 (0%) | 3 (0%) | 6 (0.01%) | 5 (0.01%) | 5 (0.01%) | 33 (0.03%) | 4 (0%) | 0 (0%) | 65 (0.07%) |
| People & Blogs | 57 (0.06%) | 71 (0.07%) | 137 (0.14%) | 488 (0.49%) | 490 (0.49%) | 557 (0.56%) | 714 (0.72%) | 753 (0.76%) | 937 (0.94%) | 949 (0.96%) | 498 (0.5%) | 5651 (5.69%) |
| Pets & Animals | 2 (0%) | 3 (0%) | 3 (0%) | 4 (0%) | 1 (0%) | 1 (0%) | 0 (0%) | 0 (0%) | 4 (0%) | 1 (0%) | 0 (0%) | 19 (0.02%) |
| Science & Technology | 0 (0%) | 1 (0%) | 11 (0.01%) | 13 (0.01%) | 4 (0%) | 4 (0%) | 62 (0.06%) | 5 (0.01%) | 5 (0.01%) | 9 (0.01%) | 2 (0%) | 116 (0.12%) |
| Shows | 0 (0%) | 44 (0.04%) | 126 (0.13%) | 259 (0.26%) | 550 (0.55%) | 224 (0.23%) | 47 (0.05%) | 2 (0%) | 1 (0%) | 1 (0%) | 0 (0%) | 1 254 (1.26%) |
| Sports | 12 (0.01%) | 17 (0.02%) | 14 (0.01%) | 106 (0.11%) | 66 (0.07%) | 78 (0.08%) | 165 (0.17%) | 217 (0.22%) | 174 (0.18%) | 97 (0.1%) | 39 (0.04%) | 985 (0.99%) |
| Travel & Events | 2 (0%) | 6 (0.01%) | 27 (0.03%) | 18 (0.02%) | 14 (0.01%) | 7 (0.01%) | 117 (0.12%) | 122 (0.12%) | 161 (0.16%) | 110 (0.11%) | 44 (0.04%) | 628 (0.63%) |
| **Total** | **428 (0.43%)** | **879 (0.88%)** | **2 039 (2.05%)** | **4 996 (5.03%)** | **10 212 (10.28%)** | **13 355 (13.44%)** | **13 482 (13.57%)** | **12 443 (12.53%)** | **13 199 (13.29%)** | **17 817 (17.94%)** | **10 484 (10.55%)** | **99 334 (100%)** |

### 3.2.3 Determining the size of the *NottDeuYTSch* corpus

As researchers on corpus linguistics have observed (e.g. Baker 2010), the appropriate size of a corpus varies depending on the features that are to be analysed. From a purely statistical standpoint, a chi-square test requires an expected value of at least five occurrences of a linguistic feature to successfully run the test. If the frequency of this feature occurs once every 10,000 tokens, then the corpus must contain at least 50,000 tokens. The *NottDeuYTSch* corpus is intended to be large enough to analyse lexical, orthographical, morphological, and syntactic features, the last of which requires a corpus size of at least one million tokens, according to Baker (2010: 95–96). Furthermore, I also used other similar corpora as a benchmark. The *NottDeuYTSch* corpus was built for purposes similar to the goals of other corpora, such as the *Dortmunder Chatkorpus*, a collection of online chat dialogues from the 2000s. Based on what other corpora have been used to research, I came to the conclusion that a corpus of at least ten million tokens (with at least one million tokens in the largest video categories and years) was necessary to provide enough data to be representative of young people's language on *YouTube*. This should also be large enough to provide a suitable number of features for grammatical and morphosyntactic analysis, as well as offer the opportunity for longitudinal examination over the ten-year period covered by the corpus.

The total number of comments under the 99,334 videos in the pre-corpus database was over 150 million, which equates to roughly 1.5bn tokens. The distribution of these comments per video category and year can be seen in Table 3.5, with both the raw numbers and percentages. This amount of data would take too long to process and analyse within the scope of a PhD project, so, as noted above, I used stratified random sampling of the pre-corpus database based on the proportions of videos under each video category and year. The smallest acceptable size for the corpus was based on the number of videos that would contribute at least 1,000 comments in the smallest category (in this case 'Pets & Animals'), which would also provide a minimum of 10,000 tokens per category. This number of tokens was sufficient for the analyses planned for the project, as well as enabling possible future inter-categorical

research, i.e. genre analyses. However, it might have been too small for a quantitative intra-categorical examination of specific aspects of the language of comments under videos of 'Pets & Animals' as a standalone corpus, as per guidelines set out by Kennedy (1998: 68) and Baker (2018: 169–71) that advise on the number of tokens necessary for corpus linguistic analysis. Using the percentages from Table 3.5, I therefore scaled the corpus down from 150 million comments to find the number of comments needed per video category and year when the total number of comments in the 'Pets & Animals' category was equal to 1,000 comments. Based on this figure, the *NottDeuYTSch* corpus should therefore have 4.8 million comments with an approximate token count of 50 million (based on my previous research on average number of tokens per *YouTube* comment), as shown in Table 3.6.

There were also a few other parameters for selecting the videos for the *NottDeuYTSch* corpus:

1. Only videos with over 100 comments were selected. My previous research on *YouTube* suggests that the average comment contains just over 10 tokens, so selecting videos with over 100 comments, should ensure that every video contributes (on average) over 1,000 words. This 1,000-word guidelines is proposed by Biber (1993: 249–52) in his work on representativeness in corpus design. A 1,000-word minimum sample size helps "to reliably represent the distributions of linguistic features" (Biber 1993: 252), such as word types. The 100-comment minimum was also applied to each year and video category.
2. Videos must be published between July 2008 and October 2018. This ensured that all videos and comments were created after *YouTube* launched the localised version of the website for Germany on 8th November 2007, which had the effect of promoting German-language content to German speakers.

Table 3.5: Number and percentage of comments under videos in each category and year in the pre-corpus database

| Category/Year | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Autos & Vehicles | 0 (0%) | 14 (0%) | 1 842 (0%) | 365 (0%) | 255 (0%) | 0 (0%) | 1 352 (0%) | 7 196 (0%) | 127 060 (0.08%) | 82 478 (0.05%) | 21 995 (0.01%) | **242 557 (0.16%)** |
| Comedy | 15 163 (0.01%) | 111 630 (0.07%) | 134 094 (0.09%) | 162 658 (0.11%) | 333 644 (0.22%) | 458 033 (0.3%) | 846 737 (0.56%) | 495 544 (0.33%) | 561 924 (0.37%) | 1 192 419 (0.79%) | 339 591 (0.23%) | **4 651 437 (3.08%)** |
| Education | 1 470 (0%) | 2 864 (0%) | 7 571 (0.01%) | 8 765 (0.01%) | 3 618 (0%) | 31 741 (0.02%) | 43 479 (0.03%) | 165 323 (0.11%) | 1 239 183 (0.82%) | 0 (0%) | 0 (0%) | **1 504 014 (1%)** |
| Entertainment | 34 112 (0.02%) | 170 202 (0.11%) | 535 573 (0.36%) | 1 124 418 (0.75%) | 1 667 487 (1.11%) | 3 736 371 (2.48%) | 3 505 933 (2.32%) | 5 280 195 (3.5%) | 17 027 527 (11.29%) | 15 199 899 (10.08%) | 7 630 318 (5.06%) | **55 912 035 (37.07%)** |
| Film & Animation | 1 657 (0%) | 5 219 (0%) | 136 903 (0.09%) | 136 150 (0.09%) | 136 050 (0.09%) | 380 483 (0.25%) | 115 764 (0.08%) | 47 557 (0.03%) | 140 517 (0.09%) | 132 059 (0.09%) | 61 995 (0.04%) | **1 294 354 (0.86%)** |
| Gaming | 304 (0%) | 2 536 (0%) | 7 432 (0%) | 223 843 (0.15%) | 1 941 302 (1.29%) | 4 327 855 (2.87%) | 3 697 528 (2.45%) | 3 960 332 (2.63%) | 5 135 070 (3.4%) | 2 325 390 (1.54%) | 1 017 213 (0.67%) | **22 638 805 (15.01%)** |
| Howto & Style | 8 947 (0.01%) | 74 260 (0.05%) | 172 709 (0.11%) | 169 145 (0.11%) | 408 209 (0.27%) | 959 509 (0.64%) | 2 535 610 (1.68%) | 3 784 621 (2.51%) | 17 777 857 (11.79%) | 15 758 660 (10.45%) | 3 154 952 (2.09%) | **44 804 479 (29.7%)** |
| Music | 37 869 (0.03%) | 30 781 (0.02%) | 29 608 (0.02%) | 39 139 (0.03%) | 73 264 (0.05%) | 58 916 (0.04%) | 54 740 (0.04%) | 122 398 (0.08%) | 152 120 (0.1%) | 1 494 841 (0.99%) | 232 238 (0.15%) | **2 325 914 (1.54%)** |
| News & Politics | 0 (0%) | 5 967 (0%) | 114 871 (0.08%) | 10 653 (0.01%) | 4 284 (0%) | 31 989 (0.02%) | 7 710 (0.01%) | 1 640 (0%) | 102 824 (0.07%) | 31 365 (0.02%) | 23 956 (0.02%) | **335 259 (0.22%)** |
| Nonprofits & Activism | 0 (0%) | 93 (0%) | 1 276 (0%) | 654 (0%) | 2 850 (0%) | 8 155 (0.01%) | 1 051 (0%) | 873 (0%) | 57 433 (0.04%) | 912 (0%) | 0 (0%) | **73 297 (0.05%)** |
| People & Blogs | 10 851 (0.01%) | 31 074 (0.02%) | 46 083 (0.03%) | 204 005 (0.14%) | 575 812 (0.38%) | 763 965 (0.51%) | 983 598 (0.65%) | 1 664 082 (1.1%) | 3 818 017 (2.53%) | 3 238 900 (2.15%) | 562 005 (0.37%) | **11 898 392 (7.89%)** |
| Pets & Animals | 884 (0%) | 2 128 (0%) | 1 054 (0%) | 1 421 (0%) | 1 211 (0%) | 461 (0%) | 0 (0%) | 0 (0%) | 22 735 (0.02%) | 1 229 (0%) | 0 (0%) | **31 123 (0.02%)** |
| Science & Technology | 0 (0%) | 22 (0%) | 4 561 (0%) | 1 063 (0%) | 197 (0%) | 2 634 (0%) | 4 442 (0%) | 481 (0%) | 43 320 (0.03%) | 34 660 (0.02%) | 1 712 (0%) | **93 092 (0.06%)** |
| Shows | 0 (0%) | 100 618 (0.07%) | 288 625 (0.19%) | 332 757 (0.22%) | 1 274 231 (0.84%) | 1 043 988 (0.69%) | 34 214 (0.02%) | 780 (0%) | 711 (0%) | 2 955 (0%) | 0 (0%) | **3 078 879 (2.04%)** |
| Sports | 2 438 (0%) | 4 121 (0%) | 4 426 (0%) | 34 879 (0.02%) | 24 975 (0.02%) | 50 957 (0.03%) | 63 024 (0.04%) | 82 659 (0.05%) | 229 866 (0.15%) | 349 879 (0.23%) | 51 041 (0.03%) | **898 265 (0.6%)** |
| Travel & Events | 27 (0%) | 937 (0%) | 5 226 (0%) | 6 992 (0%) | 7 891 (0.01%) | 13 848 (0.01%) | 223 204 (0.15%) | 136 759 (0.09%) | 436 054 (0.29%) | 166 733 (0.11%) | 59 317 (0.04%) | **1 056 988 (0.7%)** |
| **Total** | **113 722 (0.08%)** | **542 466 (0.36%)** | **1 491 854 (0.99%)** | **2 456 907 (1.63%)** | **6 455 280 (4.28%)** | **11 868 905 (7.87%)** | **12 118 386 (8.03%)** | **15 750 440 (10.44%)** | **46 872 218 (31.07%)** | **40 012 379 (26.53%)** | **13 156 333 (8.72%)** | **150 838 890 (100%)** |

Table 3.6: Projected number and percentage of comments under videos in each category and year in the *NottDeuYTSch* corpus

| Category/Year | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Autos & Vehicles | 0 (0%) | 0 (0%) | 59 (0%) | 12 (0%) | 8 (0%) | 0 (0%) | 43 (0%) | 231 (0%) | 4 083 (0.08%) | 2 650 (0.05%) | 707 (0.01%) | **7 793 (0.16%)** |
| Comedy | 487 (0.01%) | 3 587 (0.07%) | 4 309 (0.09%) | 5 226 (0.11%) | 10 720 (0.22%) | 14 717 (0.3%) | 27 206 (0.56%) | 15 922 (0.33%) | 18 055 (0.37%) | 38 313 (0.79%) | 10 911 (0.23%) | **149 453 (3.08%)** |
| Education | 47 (0%) | 92 (0%) | 243 (0.01%) | 282 (0.01%) | 116 (0%) | 1 020 (0.02%) | 1 397 (0.03%) | 5 312 (0.11%) | 39 816 (0.82%) | 0 (0%) | 0 (0%) | **48 325 (1%)** |
| Entertainment | 1 096 (0.02%) | 5 469 (0.11%) | 17 208 (0.36%) | 36 128 (0.75%) | 53 577 (1.11%) | 120 052 (2.48%) | 112 648 (2.32%) | 169 656 (3.5%) | 547 104 (11.29%) | 488 382 (10.08%) | 245 167 (5.06%) | **1 796 486 (37.07%)** |
| Film & Animation | 53 (0%) | 168 (0%) | 4 399 (0.09%) | 4 375 (0.09%) | 4 371 (0.09%) | 12 225 (0.25%) | 3 720 (0.08%) | 1 528 (0.03%) | 4 515 (0.09%) | 4 243 (0.09%) | 1 992 (0.04%) | **41 588 (0.86%)** |
| Gaming | 10 (0%) | 81 (0%) | 239 (0%) | 7 192 (0.15%) | 62 375 (1.29%) | 139 056 (2.87%) | 118 804 (2.45%) | 127 248 (2.63%) | 164 993 (3.4%) | 74 716 (1.54%) | 32 684 (0.67%) | **727 398 (15.01%)** |
| Howto & Style | 287 (0.01%) | 2 386 (0.05%) | 5 549 (0.11%) | 5 435 (0.11%) | 13 116 (0.27%) | 30 830 (0.64%) | 81 471 (1.68%) | 121 602 (2.51%) | 571 213 (11.79%) | 506 335 (10.45%) | 101 370 (2.09%) | **1 439 594 (29.7%)** |
| Music | 1 217 (0.03%) | 989 (0.02%) | 951 (0.02%) | 1 258 (0.03%) | 2 354 (0.05%) | 1 893 (0.04%) | 1 759 (0.04%) | 3 933 (0.08%) | 4 888 (0.1%) | 48 030 (0.99%) | 7 462 (0.15%) | **74 733 (1.54%)** |
| News & Politics | 0 (0%) | 192 (0%) | 3 691 (0.08%) | 342 (0.01%) | 138 (0%) | 1 028 (0.02%) | 248 (0.01%) | 53 (0%) | 3 304 (0.07%) | 1 008 (0.02%) | 770 (0.02%) | **10 772 (0.22%)** |
| Nonprofits & Activism | 0 (0%) | 3 (0%) | 41 (0%) | 21 (0%) | 92 (0%) | 262 (0.01%) | 34 (0%) | 28 (0%) | 1 845 (0.04%) | 29 (0%) | 0 (0%) | **2 355 (0.05%)** |
| People & Blogs | 349 (0.01%) | 998 (0.02%) | 1 481 (0.03%) | 6 555 (0.14%) | 18 501 (0.38%) | 24 547 (0.51%) | 31 604 (0.65%) | 53 468 (1.1%) | 122 675 (2.53%) | 104 068 (2.15%) | 18 058 (0.37%) | **382 302 (7.89%)** |
| Pets & Animals | 28 (0%) | 68 (0%) | 34 (0%) | 46 (0%) | 39 (0%) | 15 (0%) | 0 (0%) | 0 (0%) | 730 (0.02%) | 39 (0%) | 0 (0%) | **1 000 (0.02%)** |
| Science & Technology | 0 (0%) | 1 (0%) | 147 (0%) | 34 (0%) | 6 (0%) | 85 (0%) | 143 (0%) | 15 (0%) | 1 392 (0.03%) | 1 114 (0.02%) | 55 (0%) | **2 991 (0.06%)** |
| Shows | 0 (0%) | 3 233 (0.07%) | 9 274 (0.19%) | 10 692 (0.22%) | 40 942 (0.84%) | 33 544 (0.69%) | 1 099 (0.02%) | 25 (0%) | 23 (0%) | 95 (0%) | 0 (0%) | **98 926 (2.04%)** |
| Sports | 78 (0%) | 132 (0%) | 142 (0%) | 1 121 (0.02%) | 802 (0.02%) | 1 637 (0.03%) | 2 025 (0.04%) | 2 656 (0.05%) | 7 386 (0.15%) | 11 242 (0.23%) | 1 640 (0.03%) | **28 862 (0.6%)** |
| Travel & Events | 1 (0%) | 30 (0%) | 168 (0%) | 225 (0%) | 254 (0.01%) | 445 (0.01%) | 7 172 (0.15%) | 4 394 (0.09%) | 14 011 (0.29%) | 5 357 (0.11%) | 1 906 (0.04%) | **33 962 (0.7%)** |
| **Total** | **3 654 (0.08%)** | **17 430 (0.36%)** | **47 934 (0.99%)** | **78 942 (1.63%)** | **207 412 (4.28%)** | **381 355 (7.87%)** | **389 371 (8.03%)** | **506 071 (10.44%)** | **1 506 032 (31.07%)** | **1 285 621 (26.53%)** | **422 721 (8.72%)** | **4 846 541 (100%)** |

As the corpus was built by collecting all the comments from videos based on their video category and year of publication, it was necessary to calculate how many videos to collect from each intersection to provide the requisite number of comments. One potential starting point would be to select a value between the projected total number of comments (4,846,541) divided by the median number of comments per video (272) and the projected total number of comments divided by the mean number of comments per video (1,519). This indicates that selecting between 65 and 365 videos would provide the necessary amount of comments to build the *NottDeuYTSch* corpus.

However, if the 65 most commented videos were initially selected and videos were then sequentially added until the projected number of comments was reached, this would produce a corpus that would not be suitable to investigate changes in language over time, which is central to my main research question, namely, what can we learn from a corpus-driven, longitudinal approach to digital youth language? Answering these questions requires including videos from as many intersections of video category and year as possible. As shown in Table 3.6, the percentage value of many of these intersections is very low, which means that a higher overall number of videos has to be selected in order to include those less popular intersections. For example, if the corpus were to consist of 365 videos, all intersections with values of less than 0.27% would not be selected, meaning that the videos would be selected from only 39 of 176 intersections, such as 'Entertainment, 2016'. This would not have been conducive to addressing my research questions, and so the selection process was adapted to account for this. There was the possibility of including only a selection of comments from at least one video in each intersection to reach the calculated proportions. However, this would cause some methodological problems. By randomly selecting comments, potential future research questions on conversational analysis would be compromised: therefore, comments selected under a video needed to include conversational threads in their entirety. To ensure that at least one video was selected from the greatest possible number of intersections (133 out of 176, excluding the intersections where no videos or one video with very few comments, i.e. fewer than 100, were published), 0.49% was

added to all percentage values and then rounded to the nearest integer to provide the number of videos selected from each intersection. Table 3.7 shows the adjusted proportions of the number of videos that comprise the *NottDeuYTSch* corpus.

Table 3.7: Number of videos selected for the *NottDeuYTSch* corpus per category and year

| Category/Year | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Autos & Vehicles | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | **3 (1.01%)** |
| Comedy | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | **11 (3.72%)** |
| Education | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 0 (0%) | 0 (0%) | **7 (2.36%)** |
| Entertainment | 1 (0.34%) | 1 (0.34%) | 2 (0.68%) | 4 (1.35%) | 4 (1.35%) | 4 (1.35%) | 6 (2.03%) | 7 (2.36%) | 12 (4.05%) | 24 (8.11%) | 14 (4.73%) | **79 (26.69%)** |
| Film & Animation | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | **10 (3.38%)** |
| Gaming | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 4 (1.35%) | 12 (4.05%) | 19 (6.42%) | 16 (5.41%) | 13 (4.39%) | 9 (3.04%) | 6 (2.03%) | 5 (1.69%) | **86 (29.05%)** |
| Howto & Style | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 2 (0.68%) | 3 (1.01%) | 3 (1.01%) | 3 (1.01%) | 3 (1.01%) | 4 (1.35%) | 2 (0.68%) | **24 (8.11%)** |
| Music | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | **11 (3.72%)** |
| News & Politics | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | **10 (3.38%)** |
| Nonprofits & Activism | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 0 (0%) | 0 (0%) | **4 (1.35%)** |
| People & Blogs | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 2 (0.68%) | 2 (0.68%) | 2 (0.68%) | 2 (0.68%) | 2 (0.68%) | 1 (0.34%) | **16 (5.41%)** |
| Pets & Animals | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 1 (0.34%) | 0 (0%) | 0 (0%) | **1 (0.34%)** |
| Science & Technology | 0 (0%) | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 0 (0%) | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 0 (0%) | **6 (2.03%)** |
| Shows | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 2 (0.68%) | 1 (0.34%) | 1 (0.34%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | **7 (2.36%)** |
| Sports | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | **11 (3.72%)** |
| Travel & Events | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | **10 (3.38%)** |
| **Total** | **6 (2.03%)** | **12 (4.05%)** | **14 (4.73%)** | **19 (6.42%)** | **28 (9.46%)** | **36 (12.16%)** | **37 (12.5%)** | **34 (11.49%)** | **37 (12.5%)** | **44 (14.86%)** | **29 (9.8%)** | **296 (100%)** |

When the video category/year intersection stipulated that only one video was to be included, the video with the number of comments closest to the required number was selected from the pre-corpus database. For category/year intersections that required more than one video, the following steps were taken:

1. A table was created that stipulated how many videos from each channel were to be selected and the total number of comments the videos from each channel should contain. This was based on the total number of comments from the videos uploaded in that intersection by each channel. For example, in the 'Gaming, 2012' category/year intersection, I needed twelve videos that contained 62,375 comments, as per Table 3.7 above.
2. Then the proportions of videos and comments for each of the channels with respect to the whole category/year intersection were calculated. The proportion of comments posted under videos uploaded by a channel determined how many videos from each channel were selected. In the 'Gaming, 2012' example listed in Table 3.8, there were fifteen *YouTube* channels that uploaded videos to this category in this calendar year. Of the 1.9 million comments in this intersection, 66.84% (1.3 million) were posted under videos uploaded by PietSmiet, which meant that a similar proportion of the twelve videos were to be selected from this channel, in this case, eight videos with a total of 41,693 comments. The videos could have been selected in proportion to the number of videos, but this would have led to less variety in channel selection, as in some categories there were a few very prolific channels that produced significantly more videos but did not necessarily produce a similar proportion of comments. This discrepancy can be seen in Table 3.9, where the channel E IL o T IR i X ™ only uploaded 0.83% of the videos in the 'Gaming, 2012' intersection, but was responsible for 14.35% of the comments.
3. When selecting multiple videos from a single channel, e.g. the eight videos from PietSmiet, a mathematical method based on the Golden Ratio (Φ) was used to choose the correct number of videos with the necessary number of comments.[7] This ensured a wide variety of videos

[7] Φ is roughly equal to 1.618 and refers to a partitioning a whole into successively smaller seg-

Table 3.8: 'Gaming, 2012' Intersection showing totals and proportions of videos and comments uploaded and posted to each channel

| **Channel Name** | **Total Videos** | **Total Comments** | **% Videos** | **% Comments** |
|---|---|---|---|---|
| ConCrafter LUCA | 252 | 75 492 | 4.28 | 3.89 |
| DASDING | 3 | 33 | 0.05 | 0.00 |
| DasMirkoo | 1 | 243 | 0.02 | 0.01 |
| Diejungs | 8 | 26 806 | 0.14 | 1.38 |
| E IL o T IR i X ™ | 49 | 278 578 | 0.83 | 14.35 |
| Felix von der Laden | 395 | 33 758 | 6.71 | 1.74 |
| Flyinguwe | 1 | 484 | 0.02 | 0.02 |
| HeyMoritz | 2 | 1 485 | 0.03 | 0.08 |
| inscope21 | 11 | 2 824 | 0.19 | 0.15 |
| KsFreakWhatElse | 1 | 1 579 | 0.02 | 0.08 |
| Maniac | 1 | 173 | 0.02 | 0.01 |
| Paluten | 150 | 18 403 | 2.55 | 0.95 |
| PietSmiet | 4 330 | 1 297 566 | 73.55 | 66.84 |
| Sarazar | 610 | 19 2269 | 10.36 | 9.90 |
| ungespielt | 73 | 11 609 | 1.24 | 0.60 |
| **Total** | **5 887** | **1 941 302** | **100.00** | **100.00** |

Table 3.9: 'Gaming, 2012' Intersection showing how many videos and comments are to be selected from each channel

| **Channel Name** | **# Videos to select per % videos** | **# Videos to select per % comments** | **# Comments per videos per channel** |
|---|---|---|---|
| ConCrafter LUCA | 0.5 | 0.5 | 2 425 |
| DASDING | 0.0 | 0.0 | 1 |
| DasMirkoo | 0.0 | 0.0 | 8 |
| Diejungs | 0.0 | 0.2 | 861 |
| E IL o T IR i X ™ | 0.1 | 1.7 | 8 951 |
| Felix von der Laden | 0.8 | 0.2 | 1 084 |
| Flyinguwe | 0.0 | 0.0 | 15 |
| HeyMoritz | 0.0 | 0.0 | 47 |
| inscope21 | 0.0 | 0.0 | 91 |
| KsFreakWhatElse | 0.0 | 0.0 | 51 |
| Maniac | 0.0 | 0.0 | 5 |
| Paluten | 0.3 | 0.1 | 592 |
| PietSmiet | 8.8 | 8.0 | 41 693 |
| Sarazar | 1.2 | 1.2 | 6 177 |
| Ungespielt | 0.1 | 0.1 | 373 |
| **Total** | **11.8** | **12.0** | **62 374** |

would be selected, from videos that attracted a lot of comments, to those with fewer comments. In doing so, it was hoped to encompass a wide amount of language variation, with videos with fewer comments potentially containing different linguistic features. The following process was carried out to select the videos:

i) The values determined by Φ are divided by their sum to convert them to a percentage value. For example, to select three videos, the first three values Φ (0.618, 0.382, and 0.236) are added up (1.236) and each divided by that number to give the percentage values of 50%, 31%, and 19%.

ii) The new set of values are then multiplied by the total number of comments needed from the channel (e.g. 41,693 comments for PietSmiet in 'Gaming, 2012'). This provides the number of videos needed with a target number of comments.

iii) The videos with the number of comments closest to these targets are then selected. Table 3.10 shows the videos selected from the PietSmiet channel, Φ values and the number of comments under each video.

ments, in which a segment is equal to the sum of the next two smaller segments (Markowsky 1992). So, the first video is assigned the value 0.618, the second one 0.382, the third one 0.236, and so on.

Table 3.10: Eight selected videos from the PietSmiet channel in the 'Gaming, 2012' intersection, showing the F values, target number of comments, and actual number of extracted comments

| Channel | Name of Video | Upload Year | Original Φ value | Standardised Φ value | Target # Comments | Actual # Comments |
|---|---|---|---|---|---|---|
| PietSmiet | Pietsmiet's Weihnachtskalender 2012 - 2. Advent mit Hardi | 2012 | 0.618 | 0.390 | 16 277 | 15 536 |
| PietSmiet | Pietsmiet's Weihnachtskalender 2012 - 3. Advent mit Sep | 2012 | 0.382 | 0.241 | 10 061 | 10 760 |
| PietSmiet | Pietsmiet's Weihnachtskalender 2012 - 4. Advent mit Jay | 2012 | 0.236 | 0.149 | 6 216 | 5 912 |
| PietSmiet | FROHE WEIHNACHTEN 2 aka PietSmiet's Weihnachtskalender 2012 - Heiligabend | 2012 | 0.146 | 0.092 | 3 845 | 3 684 |
| PietSmiet | Trashnight mit Br4mm3n und Hardi #023 [Deutsch/HD] - Gotham City Impostors | 2012 | 0.090 | 0.057 | 2 370 | 2 201 |
| PietSmiet | Let's Play FIFA 13 [Deutsch/HD] - PietSmiet feat. Gronkh feat. Sarazar | 2012 | 0.056 | 0.035 | 1 475 | 1 370 |
| PietSmiet | Let's Play Minecraft Adventure-Maps [Deutsch/HD] - 500 Jumps to Success #3 | 2012 | 0.034 | 0.021 | 895 | 844 |
| PietSmiet | Let's Play Together DayZ #167 [Deutsch/Full-HD] - ALTER! | 2012 | 0.021 | 0.013 | 553 | 520 |
| **Total** | | | **1.583** | **1.000** | **41 693** | **40 827** |

4. After the videos had been allocated to the larger channels, there was sometimes still one video left to assign; this was often the case when there were several smaller channels in an intersection. This remaining video was selected from the other channels based on the number of comments that were needed to be collected. In the 'Gaming, 2012' example, the video was selected from the ConCrafter LUCA channel as it had the video with the number of comments closest to 5,553. The videos and number of comments extracted from the 'Gaming, 2012' are displayed in Table 3.11.

Table 3.11: Videos selected for 'Gaming, 2012' intersection with number of comments

| **Channel Name** | **Video Title** | **Comment Count** |
|---|---|---|
| E IL o T IR i X ™ | HaLo REACH : 41+ Gefällt er dir ? ELoTRiX | 5181 |
| E IL o T IR i X ™ | BF3 : Famas Power und ne Menge Themen ELoTRiX | 2977 |
| PietSmiet | Pietsmiet's Weihnachtskalender 2012 - 2. Advent mit Hardi | 15536 |
| PietSmiet | Pietsmiet's Weihnachtskalender 2012 - 3. Advent mit Sep | 10760 |
| PietSmiet | Pietsmiet's Weihnachtskalender 2012 - 4. Advent mit Jay | 5912 |
| PietSmiet | FROHE WEIHNACHTEN 2 aka PietSmiet's Weihnachtskalender 2012 - Heiligabend | 3684 |
| PietSmiet | Trashnight mit Br4mm3n und Hardi #023 [Deutsch/HD] - Gotham City Impostors | 2201 |
| PietSmiet | Let's Play FIFA 13 [Deutsch/HD] - PietSmiet feat. Gronkh feat. Sarazar | 1370 |
| PietSmiet | Let's Play Minecraft Adventure-Maps [Deutsch/HD] - 500 Jumps to Success #3 | 844 |
| PietSmiet | Let's Play Together DayZ #167 [Deutsch/Full-HD] - ALTER! | 520 |
| Sarazar | SLENDER #001 - Horror mit Facecam [HD+] [Facecam] Let's Play Slender | 4093 |
| ConCrafter LUCA | Minecraft 1.7 TOP 5 TEXTURE PACKS! | 2430 |
| **Total** | | **55508** |

## 3.3 Statistical overview of the *NottDeuYTSch* corpus

A statistical overview of the *NottDeuYTSch* corpus is presented in Table 3.12 outlining the token count, total number of comments, and key averages of the corpus. The mean number of tokens per comment (10.72) correlates with the average found in my previous research on the language of young German-speakers on *YouTube* (Cotgrove 2017). The Type-Token Ratio of the *NottDeuYTSch* corpus (0.017) is slightly lower than that of the *DWDS-Kernkorpus* (0.021) (Geyken 2010: 1), which indicates less lexical diversity, i.e. commenters use the same words more often (Kettunen 2014: 223), but the closeness of the figures implies that young people's vocabulary in *YouTube* comments is almost as broad as that found in general written communication by adults.

Table 3.12: Statistical overview of the *NottDeuYTSch* corpus

| **Statistic** | **Value** |
|---|---|
| Number of Tokens (including emoji and emoticons) | 33 760 494 |
| Number of Tokens (only lexemes) | 32 549 462 |
| Number of Types | 567 086 |
| Type-Token Ratio (TTR) | 0.017 |
| Number of Comments | 3 149 457 |
| Number of Videos | 296 |
| YouTube Channels Represented | 63 |
| Mean Tokens per Comment | 10.72 |
| Median Tokens per Comment | 5 |
| Mean Comments per Video | 1 914 |

The extracted numbers of comments for each intersection were consistently lower than the target, as shown in Table 3.11 above, in the case of the video from ConCrafter LUCA, where the video only had 2,430 comments, which is 3,000 fewer than desired. This was a trend for most intersections, and it was discovered that the reported number of comments by the *YouTube* Application Programming Interface was different to the number of comments that it was possible to extract. Some of the differences can be explained by the videos selected having fewer comments than the target number. However, 1.7 million comments of the predicted 4.8 million were not available to download using the

*YouTube* API. The main reason for this shortfall is that the comments had been removed from *YouTube* but were still counted by the *YouTube* API.[8] Despite this shortfall, the *NottDeuYTSch* corpus, with a total of 3.1 million comments from 296 covering 10 years of data from 2008 to 2018, is still a suitable size to answer the research questions.

As demonstrated by the similarities in the percentage values in Table 3.13 and Table 3.14, the targeted proportions for the distribution of comments per video category and year were generally met, although the comment timestamps in the *NottDeuYTSch* corpus are slightly more weighted towards later years. For most videos, the bulk of the comments are posted within the first two months of the upload date. However, commenters revisit older *YouTube* videos and leave comments, such as "Who is still watching this in 2017?" under a video uploaded in 2008, which is the major contributing factor to the slight difference between the targeted and achieved proportions for the distribution of comments per video category and year. This does not pose any thorny methodological problems, as the comments are timestamped.

[8] Comments can be removed by the commenter, the channel owner, or by *YouTube* themselves, if the comment violates their community guidelines.

Table 3.13: Number of comments per video category and year in the *NottDeuYTSch* corpus

| category | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Autos & Vehicles | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 3 508 (0.11%) | 2 306 (0.07%) | 704 (0.02%) | **6 518 (0.21%)** |
| Comedy | 434 (0.01%) | 439 (0.01%) | 2 986 (0.09%) | 1 686 (0.05%) | 5 405 (0.17%) | 9 326 (0.3%) | 23 731 (0.75%) | 10 279 (0.33%) | 25 447 (0.81%) | 36 908 (1.17%) | 13 116 (0.42%) | **129 757 (4.12%)** |
| Education | 0 (0%) | 29 (0%) | 196 (0.01%) | 300 (0.01%) | 139 (0%) | 19 (0%) | 115 (0%) | 2 470 (0.08%) | 28 046 (0.89%) | 5 422 (0.17%) | 3 914 (0.12%) | **40 650 (1.29%)** |
| Entertainment | 494 (0.02%) | 2 590 (0.08%) | 7 236 (0.23%) | 13 854 (0.44%) | 22 588 (0.72%) | 70 378 (2.23%) | 39 747 (1.26%) | 81 248 (2.58%) | 340 062 (10.8%) | 348 523 (11.07%) | 203 871 (6.47%) | **1 130 591 (35.9%)** |
| Film & Animation | 0 (0%) | 86 (0%) | 3 781 (0.12%) | 3 204 (0.1%) | 3 503 (0.11%) | 1 056 (0.03%) | 4 561 (0.14%) | 2 124 (0.07%) | 4 373 (0.14%) | 4 044 (0.13%) | 2 069 (0.07%) | **28 801 (0.91%)** |
| Gaming | 0 (0%) | 24 (0%) | 118 (0%) | 883 (0.03%) | 49 482 (1.57%) | 89 822 (2.85%) | 78 498 (2.49%) | 97 392 (3.09%) | 124 231 (3.94%) | 70 262 (2.23%) | 30 754 (0.98%) | **541 466 (17.19%)** |
| Howto & Style | 5 (0%) | 380 (0.01%) | 3 585 (0.11%) | 5 119 (0.16%) | 5 188 (0.16%) | 15 998 (0.51%) | 52 969 (1.68%) | 63 920 (2.03%) | 323 636 (10.28%) | 345 654 (10.98%) | 51 115 (1.62%) | **867 569 (27.55%)** |
| Music | 0 (0%) | 93 (0%) | 1 219 (0.04%) | 849 (0.03%) | 1 610 (0.05%) | 1 723 (0.05%) | 2 227 (0.07%) | 1 210 (0.04%) | 5 446 (0.17%) | 35 710 (1.13%) | 7 436 (0.24%) | **57 523 (1.83%)** |
| News & Politics | 0 (0%) | 86 (0%) | 2 699 (0.09%) | 416 (0.01%) | 195 (0.01%) | 195 (0.01%) | 339 (0.01%) | 46 (0%) | 3 390 (0.11%) | 1 071 (0.03%) | 1 654 (0.05%) | **10 091 (0.32%)** |
| Nonprofits & Activism | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 184 (0.01%) | 18 (0%) | 68 (0%) | 1 571 (0.05%) | 142 (0%) | 34 (0%) | **2 017 (0.06%)** |
| People & Blogs | 25 (0%) | 818 (0.03%) | 566 (0.02%) | 3 405 (0.11%) | 6 290 (0.2%) | 12 501 (0.4%) | 7 258 (0.23%) | 16 762 (0.53%) | 52 608 (1.67%) | 77 513 (2.46%) | 16 871 (0.54%) | **194 617 (6.18%)** |
| Pets & Animals | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 1 053 (0.03%) | 8 (0%) | 3 (0%) | **1 064 (0.03%)** |
| Science & Technology | 0 (0%) | 0 (0%) | 32 (0%) | 54 (0%) | 20 (0%) | 6 (0%) | 12 (0%) | 51 (0%) | 986 (0.03%) | 1 101 (0.03%) | 11 (0%) | **2 273 (0.07%)** |
| Shows | 0 (0%) | 245 (0.01%) | 4 896 (0.16%) | 5 747 (0.18%) | 26 482 (0.84%) | 37 874 (1.2%) | 1 951 (0.06%) | 1 775 (0.06%) | 1 865 (0.06%) | 1 152 (0.04%) | 616 (0.02%) | **82 603 (2.62%)** |
| Sports | 11 (0%) | 56 (0%) | 170 (0.01%) | 565 (0.02%) | 948 (0.03%) | 1 520 (0.05%) | 1 801 (0.06%) | 479 (0.02%) | 5 801 (0.18%) | 11 819 (0.38%) | 2 368 (0.08%) | **25 538 (0.81%)** |
| Travel & Events | 0 (0%) | 26 (0%) | 138 (0%) | 198 (0.01%) | 204 (0.01%) | 384 (0.01%) | 2 079 (0.07%) | 4 956 (0.16%) | 12 149 (0.39%) | 6 272 (0.2%) | 1 973 (0.06%) | **28 379 (0.9%)** |
| **Total** | **969 (0.03%)** | **4 872 (0.15%)** | **27 622 (0.88%)** | **36 280 (1.15%)** | **122 054 (3.88%)** | **240 986 (7.65%)** | **215 306 (6.84%)** | **282 780 (8.98%)** | **934 172 (29.66%)** | **947 907 (30.1%)** | **336 509 (10.68%)** | **3 149 457 (100%)** |

Table 3.14: Projected number and percentage of comments under videos in each category and year in the *NottDeuYTSch* corpus

| Category/Year | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Autos & Vehicles | 0 (0%) | 0 (0%) | 59 (0%) | 12 (0%) | 8 (0%) | 0 (0%) | 43 (0%) | 231 (0%) | 4 083 (0.08%) | 2 650 (0.05%) | 707 (0.01%) | **7 793 (0.16%)** |
| Comedy | 487 (0.01%) | 3 587 (0.07%) | 4 309 (0.09%) | 5 226 (0.11%) | 10 720 (0.22%) | 14 717 (0.3%) | 27 206 (0.56%) | 15 922 (0.33%) | 18 055 (0.37%) | 38 313 (0.79%) | 10 911 (0.23%) | **149 453 (3.08%)** |
| Education | 47 (0%) | 92 (0%) | 243 (0.01%) | 282 (0.01%) | 116 (0%) | 1 020 (0.02%) | 1 397 (0.03%) | 5 312 (0.11%) | 39 816 (0.82%) | 0 (0%) | 0 (0%) | **48 325 (1%)** |
| Entertainment | 1 096 (0.02%) | 5 469 (0.11%) | 17 208 (0.36%) | 36 128 (0.75%) | 53 577 (1.11%) | 120 052 (2.48%) | 112 648 (2.32%) | 169 656 (3.5%) | 547 104 (11.29%) | 488 382 (10.08%) | 245 167 (5.06%) | **1 796 486 (37.07%)** |
| Film & Animation | 53 (0%) | 168 (0%) | 4 399 (0.09%) | 4 375 (0.09%) | 4 371 (0.09%) | 12 225 (0.25%) | 3 720 (0.08%) | 1 528 (0.03%) | 4 515 (0.09%) | 4 243 (0.09%) | 1 992 (0.04%) | **41 588 (0.86%)** |
| Gaming | 10 (0%) | 81 (0%) | 239 (0%) | 7 192 (0.15%) | 62 375 (1.29%) | 139 056 (2.87%) | 118 804 (2.45%) | 127 248 (2.63%) | 164 993 (3.4%) | 74 716 (1.54%) | 32 684 (0.67%) | **727 398 (15.01%)** |
| Howto & Style | 287 (0.01%) | 2 386 (0.05%) | 5 549 (0.11%) | 5 435 (0.11%) | 13 116 (0.27%) | 30 830 (0.64%) | 81 471 (1.68%) | 121 602 (2.51%) | 571 213 (11.79%) | 506 335 (10.45%) | 101 370 (2.09%) | **1 439 594 (29.7%)** |
| Music | 1 217 (0.03%) | 989 (0.02%) | 951 (0.02%) | 1 258 (0.03%) | 2 354 (0.05%) | 1 893 (0.04%) | 1 759 (0.04%) | 3 933 (0.08%) | 4 888 (0.1%) | 48 030 (0.99%) | 7 462 (0.15%) | **74 733 (1.54%)** |
| News & Politics | 0 (0%) | 192 (0%) | 3 691 (0.08%) | 342 (0.01%) | 138 (0%) | 1 028 (0.02%) | 248 (0.01%) | 53 (0%) | 3 304 (0.07%) | 1 008 (0.02%) | 770 (0.02%) | **10 772 (0.22%)** |
| Nonprofits & Activism | 0 (0%) | 3 (0%) | 41 (0%) | 21 (0%) | 92 (0%) | 262 (0.01%) | 34 (0%) | 28 (0%) | 1 845 (0.04%) | 29 (0%) | 0 (0%) | **2 355 (0.05%)** |
| People & Blogs | 349 (0.01%) | 998 (0.02%) | 1 481 (0.03%) | 6 555 (0.14%) | 18 501 (0.38%) | 24 547 (0.51%) | 31 604 (0.65%) | 53 468 (1.1%) | 122 675 (2.53%) | 104 068 (2.15%) | 18 058 (0.37%) | **382 302 (7.89%)** |
| Pets & Animals | 28 (0%) | 68 (0%) | 34 (0%) | 46 (0%) | 39 (0%) | 15 (0%) | 0 (0%) | 0 (0%) | 730 (0.02%) | 39 (0%) | 0 (0%) | **1 000 (0.02%)** |
| Science & Technology | 0 (0%) | 1 (0%) | 147 (0%) | 34 (0%) | 6 (0%) | 85 (0%) | 143 (0%) | 15 (0%) | 1 392 (0.03%) | 1 114 (0.02%) | 55 (0%) | **2 991 (0.06%)** |
| Shows | 0 (0%) | 3 233 (0.07%) | 9 274 (0.19%) | 10 692 (0.22%) | 40 942 (0.84%) | 33 544 (0.69%) | 1 099 (0.02%) | 25 (0%) | 23 (0%) | 95 (0%) | 0 (0%) | **98 926 (2.04%)** |
| Sports | 78 (0%) | 132 (0%) | 142 (0%) | 1 121 (0.02%) | 802 (0.02%) | 1 637 (0.03%) | 2 025 (0.04%) | 2 656 (0.05%) | 7 386 (0.15%) | 11 242 (0.23%) | 1 640 (0.03%) | **28 862 (0.6%)** |
| Travel & Events | 1 (0%) | 30 (0%) | 168 (0%) | 225 (0%) | 254 (0.01%) | 445 (0.01%) | 7 172 (0.15%) | 4 394 (0.09%) | 14 011 (0.29%) | 5 357 (0.11%) | 1 906 (0.04%) | **33 962 (0.7%)** |
| **Total** | **3 654 (0.08%)** | **17 430 (0.36%)** | **47 934 (0.99%)** | **78 942 (1.63%)** | **207 412 (4.28%)** | **381 355 (7.87%)** | **389 371 (8.03%)** | **506 071 (10.44%)** | **1 506 032 (31.07%)** | **1 285 621 (26.53%)** | **422 721 (8.72%)** | **4 846 541 (100%)** |

## 3.4 Applications of the *NottDeuYTSch* corpus

The *NottDeuYTSch* corpus is one of the first large corpora of linguistic data containing language written specifically by young German-speakers in *YouTube* comments, an important and popular site of youth culture and discourse. The corpus is thus a significant contribution to corpora of online data, complementing existing corpora, which focus on other areas of online language, such as the *MoCoDa2* corpus of WhatsApp messages (Beißwenger et al. 2020), the *DiDi* corpus of Facebook texts (Glaznieks & Frey 2020), and the *IBK* corpus of multiple online sources, e.g. emails, IRC chats, and Blogs (Lüngen & Kupietz 2020). The *NottDeuYTSch* corpus offers a wide range of new possibilities for study. The structured sampling of the data over the time frame of the corpus enables a wide range of longitudinal studies for lexical, orthographical, and morphosyntactic features. Videos and comments contain a wealth of metadata, which can facilitate a wide range of future research, e.g. analyses of video genres, timeframes, users, or *YouTubers*. The metadata also allow interactional and discourse analyses of interactions between commenters as it preserves the comment structure on a page, i.e. parent comments and replies.

The large size of the *NottDeuYTSch* corpus allows for considerable quantitative research, including the investigation of features that do not occur frequently, such as some syntactic constructions, as well as linguistic features specific to Digitally-Mediated Communication, such as emoji and hashtags, where a large amount of data is required for linguistic study beyond qualitative analysis. The comments in the *NottDeuYTSch* corpus are predominantly written in German (including dialect use), but there is also a significant presence of other languages, such as English, Turkish, and Russian, including linguistic elements from multiple languages within the same comment, and the corpus can also be used for potential quantitative and qualitative analyses of multilingualism. In the following chapters, three linguistic case studies are presented, which each focus on a different area of linguistics; lexis, morphosyntax, and orthography, demonstrating the wide applicability of the *NottDeuYTSch* corpus to analyse the digital writing of young people.

# 4

# *Affengeil* and *abogeil* – lexical and morphological developments in digital youth language

This chapter investigates the use of *geil* in the *NottDeuYTSch* corpus. *Geil* is perhaps the word most associated with youth language in German popular culture (see Schlobinski, Kohl, & Ludewigt 1993: 63; Eichinger 2001: 347; Mroczynski 2018: 328-329). The word *geil* is polysemous and has four senses in contemporary usage, which, as listed in the Duden (Kunkel-Razum et al. 2020), are:

1. "gierig nach geschlechtlicher Befriedigung, vom Sexualtrieb beherrscht, sexuell erregt" ('greedy for sexual gratification, dominated by the sex drive, sexually aroused').
2. "(von Pflanzen) [allzu] üppig, aber nicht sehr kräftig wachsend; wuchernd" ('(of plants) growing [too] luxuriantly, but not very vigorously; rampant').
3. "(auf etwas geil sein; -geil) auf etwas versessen" ('to be hell-bent on something, desperate for something').
4. "in begeisternder Weise schön, gut; großartig, toll" ('pleasant in an enthu-

siastic way, good, great, cool').

Use of the fourth sense, arising from a semantic shift in the latter half of the twentieth century, has frequently appeared in scholarship on *Jugendsprache* since the 1980s. This sense of *geil* has also heavily featured in the media, spawning several musical hits, for example "Geil (Geilomatick Mix)" by Bruce & Bongo (1986), where *geil* is referred to as "Germany's most successful word", reached number one in the German and Austrian charts, and "Supergeil (feat. Friedrich Liechtenstein)" by Der Tourist (2013) was used in an advert in 2014 for supermarket chain EDEKA (*Supergeil (Feat. Friedrich Liechtenstein) - EDEKA Version* (2014)). The word was also the main strapline for the consumer electronics retailer *Saturn* between 2002 and 2011, i.e. *Geiz ist geil* ('stinginess is cool'). In the *NottDeuYTSch* corpus, *geil* is the 30th most frequent lemma, occurring roughly as often as *was* ('what') and *wenn* ('if'; 'when'). However, the second sense of *geil* listed above, i.e. referring to plants, is not attested in the corpus at all, so the three remaining senses of *geil* form the basis for the analyses in this chapter.

The analyses are informed by previous scholarship on *geil*, which has been examined in Chapter 2, Section 2.2.3. Additionally, the senses of *geil* are contextualised within a historical overview in Section 4.1, alongside an overview of the current usages of the word in contemporary German. The section illustrates both the senses and the lexical categories of *geil*, using examples from the *NottDeuYTSch* corpus, as well as from the *Deutsche Referenz Korpus* (henceforth *DeReKo*) and the *Digitales Wörterbuch der deutschen Sprache Kernkorpus* (*DWDS Kernkorpus*) for historical comparison since 1900.

Section 4.2 provides a short overview of the frequency of *geil* in the *NottDeuYTSch* corpus. Section 4.3 analyses morphological developments in the three attested senses of *geil* using both qualitative and quantitative methodologies. We shall see that in addition to its high frequency, *geil* is also very productive in the *NottDeuYTSch* corpus, occurring in many compounds, derivations, inflected forms, and orthographic variations. Section 4.3.2 analyses the use of *geil* as an affixoid (whether as a 'prefixoid', i.e. a morpheme before the stem, or 'suffixoid', i.e. a morpheme after the stem), in particular,

focusing on the evaluative properties of *geil* as a suffixoid. Section 4.3.3 examines the morphological intensification of *geil* as a base lexeme through derivation and compounding. The section then focuses on the use of multiple intensifying affixes and affixoids in a single lexeme, e.g. *superhammermegageil* (lit. 'super-hammer-mega-cool', i.e. 'mega-super-duper-cool'), demonstrating that lexemes with multiple intensifiers have an internal grammatical structure, which informs how complex forms of *geil* are constructed in German youth language.

In previous scholarship, longitudinal studies of the use of *geil* in *Jugendsprache* have been rare, so Section 4.4 investigates changes in the usage of *geil* in the *NottDeuYTSch* corpus over the period 2008-2018, examining the frequencies of all three attested senses of *geil*. The analysis demonstrates a dramatic decrease in the usage of *geil* as a positive evaluation, i.e. 'cool', but a substantial increase in the usage of of *geil* to mean 'having a strong desire (towards something)'. The section examines the potential reasons for these changes, and presents evidence from the corpus for words and phrases that are being used instead by young German-speakers to express a positive evaluation.

## 4.1 *Geil* in history and scholarship

The adjective *geil* first appeared in Old High German in the 8th Century, originating from the Germanic adjective, <*gaila->, meaning "lustig, lüstern" ('joyful, lustful') (Kluge & Seebold 2002). However, in New High German, the meaning acquired sexually charged connotations, and diverged into two: "sexually lustful" (Sense 1, as shown in Example 4.1), and "luxuriant but feeble" (Sense 2), which pertained to botanical conditions, and was therefore restricted to specialised contexts. In the 1930s, *geil* acquired a third, more semantically generalised sense to indicate a strong desire towards an object, often with negative connotations, as seen in Example 4.2 from an article in the *Völkischer Beobachter*, the National Socialist-run newspaper.

**Example 4.1** (*geil* used in Sense 1. C. Jung, Psychologische Typen (1921),

DWDS-KK 1921/jung_typen_1921)
Wohnt auch ein Herz in eurem Leib und ruhet auch in eurer Seele ein Gewissen, daß ihr Solches wagt und leget also öffentlich vor Aller Augen diese bloße unverschämte **geile** Nacktheit? ('Is there also a heart in your body and a conscience in your soul that you dare to do such things and thus publicly expose this bare, shameless, **rapacious** nakedness before everyone's eyes?')

**Example 4.2** (*geil* used in Sense 3. *Völkischer Beobachter* (1930), DWDS-KK 1930/VB_Y19300301_005_0023)
Das Volk der Dichter und Denker zur Dirne dividendenhungriger Finanzhyänen herabgesunken, **machtgeilen** Journalistenhirnen willfährig, wenn man ihm schmeichelt, kritiklos, entnervt, entwertet, entwurzelt. ('The people of poets and thinkers have sunk to the level of the servant of dividend-hungry financial hyenas, compliant to **power-hungry** journalists' brains when they are flattered, uncritical, dispirited, devalued, uprooted.')

The usage of this sense of *geil* to mean "desperate for something" is mainly found in two grammatical structures: [NOUN]-*geil*, an adjectival compound, and *geil auf* followed by a noun phrase in the accusative, analogous to the use of *Lust auf* ('a desire for'):

1. [NOUN]-*geil*: *das Geld* ('money') > *geldgeil/geld-geil* ('money-grubbing')

   **Example 4.3** (*Berliner Zeitung* (2014), DeReKo B14/NOV.02212)
   […] Dass sie so **geld-geil** und oberflächlich ist. Ich sage jetzt mal nur ein Beispiel. Wir sind der erste Zirkus, der dafür bekannt war, dass er ohne Tierschau durch die Lande zieht […] ('[…] That she is so **money-grubbing** and superficial. Let me give you just one example. We are the first circus that was known for touring the country without an animal show […]')

2. *geil auf* $_{acc}$[NOUN PHRASE] > *geil auf bürgerliche Kultur*

   **Example 4.4** (*Die Zeit* (1969), DeReKo Z69/MAR.00187)

> Und weiter im Text: "Genug! Es ist höchste Zeit, daß die 'braven Studenten' und die anderen Schafe, die so **geil auf** bürgerliche Kultur sind, sich klar darüber werden, in welche Falle sie geraten. Hinaus mit dem Repräsentanten des Großkapitals und der herrschenden Ideologie! Die Kunst ist tot. Es lebe die Revolution!" ('And further in the text: "Enough! It is high time that the 'well-behaved students' and the other sheep who are so **desperate for** bourgeois culture realise what trap they are falling into. Out with the representative of big business and the ruling ideology! Art is dead. Long live the revolution!" ')

The first attestation in the *DeReKo* of *geil* to mean 'cool' (Sense 4) dates from 1979 (see Example 4.5), in a review of the social drama film, "Das Ende des Regenbogens" ('The End of the Rainbow'), where it refers to dialogue used by the main character, Jimmi, a 17-year-old sex worker in West Berlin. This attestation corresponds with Neuland's meta-analysis of 1970s and 1980s German dictionaries (Neuland 1994: 89-90), which identified that the dictionary entry for *geil* was changed to be marked as *jugendsprachlich* between 1975 and 1983, from being previously unmarked. A statistical analysis of the relative frequencies of *geil* in the *DWDS Kernkorpus* from 1900-2010 clearly illustrates the widespread adoption of this sense at this time, with a sharp increase in the 1980s (Figure 4.1).

> **Example 4.5** (*geil* used in Sense 4. Der Spiegel (1979), DeReKo S79/NOV.00175)
>
> […] Thomas Kufahl in "Das Ende des Regenbogens" ist das Gegenteil, dynamisch und aggressiv, unberechenbar wie ein verspielter Haushund: liebenswert und gefährlich, hilflos und explosiv. Er findet alles "**geil**, wa?!" oder "Scheiße!" und versteckt seinen Charme, seine Verzweiflung hinter forschem Trotz. ('[…] Thomas Kufahl in "The End of the Rainbow" is the opposite; dynamic and aggressive, unpredictable like a playful house dog: lovable and dangerous, helpless and explosive. He finds everything "**cool**, right?!" or "Shit!" and hides his charm and his desperation behind brash defiance.')

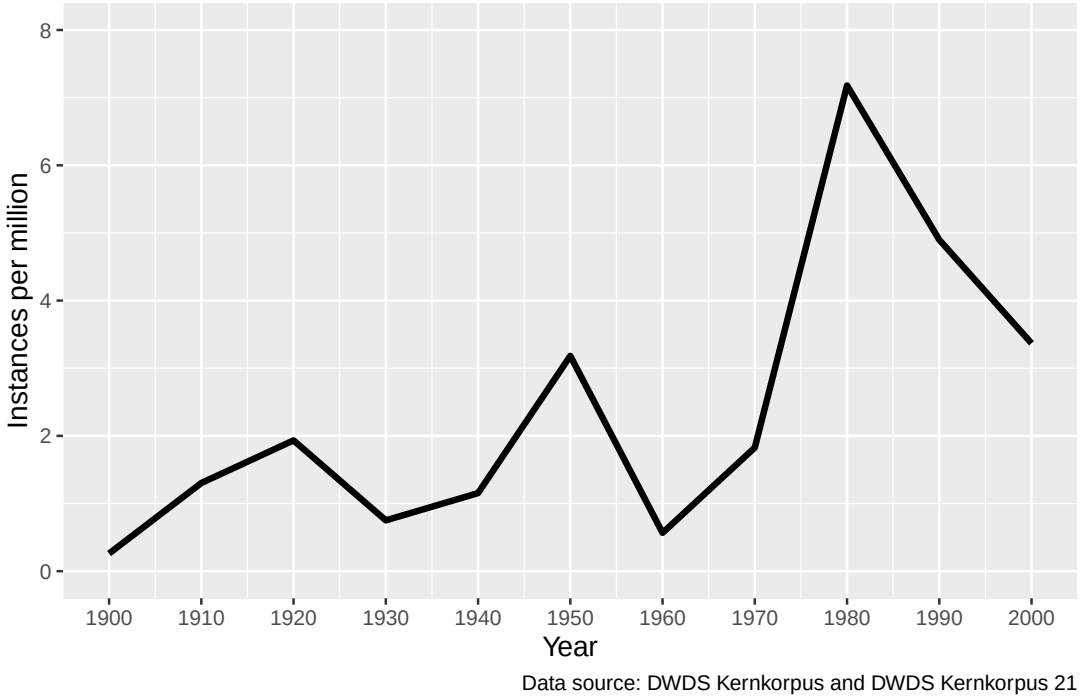



Figure 4.1: Frequency of *geil* in the *DWDS Kernkorpora* (These data do not include online sources of language)

In the *DWDS Kernkorpus* the frequency of *geil* subsequently drops off after 1980, although this is not evident in other contemporary corpora. In an analysis of three newspapers in the *IDS Zeitungskorpus* ('IDS Newspaper corpus'), Androutsopoulos (2005: 192) identified a steady increase in the frequency of the term from 1986 to 2003 (see Figure 4.2). It is this fourth sense of *geil* that has featured most frequently in studies of German youth language, and these studies are examined in the following section.

### 4.1.1 Grammatical productivity of *geil*

Using Androutsopoulos (1998) and Bahlo et al. (2019), I have identified six categories of grammatical productivity that have been characterised as typical of German youth language ('intensification', 'derivation', 'lexical innovation', 'borrowing', 'set phrases', and 'discourse markers') to demonstrate the range of uses of *geil* in the *NottDeuYTSch* corpus. While Androutsopoulos (1998) used a variety of lexemes as examples of grammatical productivity, I have used examples from the *NottDeuYTSch* corpus that all contain *geil* for each of the categories.

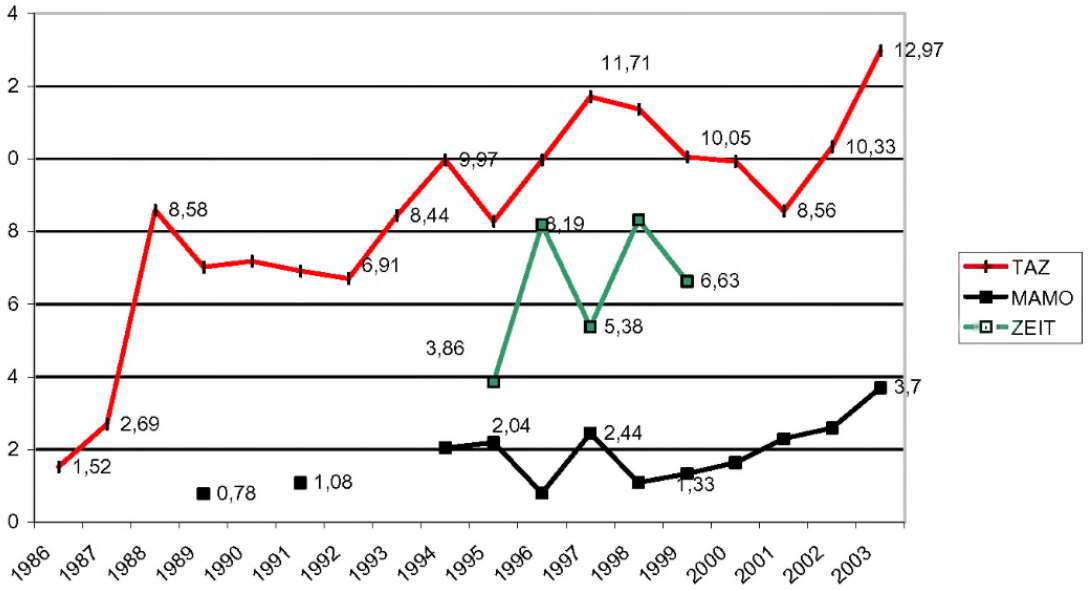



Figure 4.2: Frequency of *geil* in three newspapers (*Die Tageszeitung* - TAZ, *Mannheimer Morgen* - MAMO, and *Die Zeit* - ZEIT) from 1986-2003, taken from Androutsopoulos (2005: 192). Frequencies in ipm.

**1. Intensification** is the grammatical phenomenon of emphasising a word or phrase. Androutsopoulos (1998) outlined two main forms of intensification: compounding with affixoids (often with multiple affixoids in *Jugendsprache*), as in (a),[1] which I refer to as 'morphological intensification', and the use of particles, as in (b), which I refer to as 'syntactic intensification'.

a. Intensifying compounding: *mega* ('mega') + *ober* ('upper/above') + *Affe* ('monkey') + *geil* > *megaoberaffengeil* ('mega-above-monkey-cool', i.e. super-duper-mega cool)

**Example 4.6** (2016)

Die ersten folgen waren schon Hammer aber jetzt ist es einfach **Megaoberaffengeil** ('The first episodes were already great but now it is simply **Super-duper-mega cool**')

b. Particles: *echt* > *echt geil* ('really cool')

**Example 4.7** (2018)

[1]The semantic alteration of a base lexeme through a prefix or prefixoid without changing the word class is also referred to as 'modification': intensification is one form of this (Bahlo et al. 2019: 60).

> Eis mit pommes ist **echt geil**👌😍 ('Ice cream with chips is **really cool**👌😍')

Studies of intensification in previous research on youth language have frequently given examples containing *geil*, for example, the use of *geil* as the base lexeme in a wide range of adjectival compounds such as *saugeil* (lit. 'sow/pig-cool') (Henne 1986: 153), *ultra-*, *ober-* ('above, super'), and *supergeil* (Androutsopoulos 1998: 103-107), *affengeil* (lit. 'monkey-cool') (Androutsopoulos 2005: 179), and *megageil* (Mroczynski 2018). Intensification through the use of multiple prefixes, as shown in Example 4.6, is also briefly mentioned in work on *Jugendsprache*, e.g. *oberaffengeil* (lit. "above-monkey-cool," Schlobinski, Kohl, & Ludewigt 1993: 113), and *megahammageil* (lit. "mega-hammer-cool," Vasiljevič 2018: 921).

Schlobinski, Kohl, & Ludewigt (1993: 63–64) disputed the authenticity of some of these *geil* compounds, particularly *affengeil*, claiming that they were a construction of popular media, and only existed in the perception of *Jugendsprache* of adults. However, regardless of their possible origins as media constructions, these compounds have since been adopted by young people. The *NottDeuYTSch* corpus contains many examples of both *affengeil* and the use of multiple prefixes with *geil*, as we shall see in detail in Section 4.3.3.2.[2]

**2. Derivation** is the modification of an existing lexeme, typically through affixation, to create a new lexeme, which can change the word class, as in (c). When deriving verbs, as in (d), Androutsopoulos (1998) observed that the *ab-* and *rum-* prefixes were characteristic of German youth language, although in the *NottDeuYTSch* corpus, these do not occur with *geil*. Zero-derivation, i.e. changing the word class of a lexeme without affixation, also occurs with *geil*, as in (e).

[2] In addition to morphological and syntactic intensification, orthographic intensification is also a common strategy in digital youth language (Hilte, Vandekerckhove, & Daelemans 2019: 296). It is most often represented by the repetition of graphemes in a word, e.g. *geilllllll!* (Androutsopoulos 2003b: 186) and *di geeeilshT* (Dürscheid, Wagner, & Brommer 2010: 139). Orthographic practices such as these that can alter how the message is understood by the reader are analysed in depth in Chapter 7.

c. Adjective to noun: *geil* + *-heit* ('-ness', suffix to derive nouns) > *Geilheit* ('coolness')

**Example 4.8** (2010)
Wie immer pure **geilheit** xd ('As always pure **coolness** xd')

d. Adjective to verb: *auf* ('at/on/onto/upon') + *geil* > *aufgeilen* ('to excite')

**Example 4.9** (2013)
was geht in deren **hassaufgegeilten** minihirns ab?? ('What is going on in their **hate-fuelled** minibrains??')

e. Adjective to interjection: *geil* + *ø* > *geil* (zero-derivation)

**Example 4.10** (2012)
**Geil**, aber leider versteh ich ab 1:01 nichts mehr. Kann mir wer weiterhelfen? ('**Cool**, but unfortunately I can't understand anything from 1:01. Can somebody help me?')

**3. Lexical innovation** is the invention of new words, modification of existing words through compounding (see f), abbreviation (g), derivation, or other ludic violations of standard grammar rules (h).

f. Innovative compounding: *der Hammer* ('hammer') + *geil* > *hammergeil* ('hammer-cool', i.e. very cool)

**Example 4.11** (2012)
ich will gewinnen.. hnmm..weil hobbits einfach **hammergeil** sind. ('I want to win.. hmmm..because hobbits are simply **so cool**')

g. Abbreviations: *das Abonnement* ('subscription') > *Abo* ('sub') + *geil* > *abogeil* ('desperate for subscriptions')

**Example 4.12** (2017)
[YOUTUBER] - wie krass **Abogeil** ey :D ('[YOUTUBER] - how ridiculously **desperate for subs** eh :D')

h. Ludic innovations: *geil* + *ig* ('-y', suffix to derive adjectives) > *geilig*. This is

a superfluous derivation, as *geil* is already an adjective. As seen in Example 4.13, the formation is a portmanteau of *geil* and *heilig* ('holy'), with a sexual innuendo-laden pun on *Heiligabend* ('Christmas Eve').

**Example 4.13** (2012)
Ich brauch die Tickets für meine Freundin das wäre das perfekte Weihnachtsgeschenk!!!! Sie ist ein riesen Fan von Der Hobbit. Damit würd ich mir auch noch einen **Geilig** Abend sichern ;) ('I need the tickets for my girlfriend that would be the perfect Christmas present!!!! She is a huge fan of The Hobbit. With them I would also secure a **sexy** night ;)')

Androutsopoulos (1998: 531–32) noted some orthographic innovations in German youth language, observing that non-standard spellings may be used by young people to indicate affinity to particular music styles, such as < riddim > for < rhythm > taken from African American and Caribbean musical culture. However, he did not give examples with *geil*. Creative spellings of *geil* have been cited in other scholarship, e.g. in studies of dialect online (Aschwanden 2001: 66; Dürscheid, Wagner, & Brommer 2010: 139). As shown in Example 4.14, the morpheme < -scht >, suffixed to *geil*, represents the spoken pronunciation of the superlative in Swiss-German (conventionally spelled < -st >).

**Example 4.14** (Aschwanden (2001: 66))
Das esch de **geilscht** writer i de schwiz! ('That is the **coolest** writer in Switzerland!')

**4. Borrowing:** is the use of resources from other languages. Borrowings are often taken from areas relevant to youth culture (Androutsopoulos 1998: 544), as demonstrated in Example 4.15 where 'like' is taken from the (English) language of social media.

i. "like", from English + *geil* > *likegeil* ('desperate for likes')

**Example 4.15** (2013)
könnte daran liegen dass du **likegeil** bist ('could be because you are **desperate for likes**')

**5. Set phrases:** are the occurrence of typical collocations. For example, *geil ey* ('cool, right') was investigated by Schlobinski, Kohl, & Ludewigt (1993) and other set phrases with *geil* are reported by both Auer (2016) and Vasiljevič (2018), e.g. *wie geil ist das denn?*. Androutsopoulos (1998: 222) provided several examples of set phrases containing *geil* + [NOUN], e.g. *geile Mucke* ('cool music'). As might be expected from a collection of *YouTube* comments, variations on the collocation of *geil* and video, e.g. *geiles video* ('cool video'), appear frequently in the *NottDeuYTSch* corpus (n = 11,125, 3,532 ipmc, equivalent to the 950th most frequent word in the corpus).

> **Example 4.16** (2009)
> Wasn **geiles** Video xDDDD *rofl* ('What a **cool** video xDDD *rofl*')

**6. Discourse markers:** are the use of lexical items to connect and organise interaction. When used as a discourse marker (see zero-derivation above), *geil* or phrases containg *geil* often function as an interjection, e.g. *geil!*, *einfach geil* ('simply cool'), *wie geil* ('how cool').

> **Example 4.17** (2009)
> **wie geil**! den pulli von [YOUTUBER] hab ich auch :D ('**how cool**! I have got the same jumper as [YOUTUBER] as well :D')

The above six categories demonstrate how *geil* can be used within many linguistic constructions characteristic of youth language. Additionally, *geil* occurs in many other grammatical constructions in the *NottDeuYTSch* corpus, which are used to inform the lexical and morphological analyses carried out in this chapter. Table 4.1 lists twelve grammatical constructions adapted from earlier work by Mroczynski (2018: 331), updated using data from the *NottDeuYTSch* corpus, to demonstrate the morphological variation that can occur with *geil*, as well as the different parts of speech that can be represented, demonstrating the usefulness of *geil* to illustrate developments in German youth language.

Regardless of its debated status as either *jugendsprachlich* or *umgangssprachlich*, *geil* remains an important lexical item in German youth language itself, and

Table 4.1: Grammatical categories of *geil*, adapted from Mroczynski (2018: 331)

| **Grammatical Category** | **Example from the NottDeuYTSch corpus** |
|---|---|
| 'Geil' as a predicative adjective | das Video ist geil |
| 'Geil' as an attributive adjective | geile parodie |
| Ellipsis involving 'geil' | echt voll geil |
| Prefixoid with 'geil' | hammergeil |
| 'Geil' as a suffixoid | likegeil |
| Prefixation of 'geil' | ungeil |
| Suffixation of 'geil' | geilheit |
| 'Geil' as an infix | Conigeilcrafter |
| Circumfixation of 'geil' | aufgegeilt |
| Morphological intensification of 'geil' | geil > hammergeil > obermegahammergeil |
| Orthographical intensification of 'geil' | geil > geeeeil > geeeeeeeeeeeeeeeeeeeeeeeeeee-eiiiiiiiiiiil |
| 'Geil' as a discourse marker | geil xd |

consequently, within research on German youth language.

## 4.2 *Geil* in the *NottDeuYTSch* corpus

In the *NottDeuYTSch* corpus, *geil* is the most frequent adjective of all, appearing 4,581 times per million tokens (n = 149,102) and in 2,027 different types (i.e. all compounds, derivations, inflected forms, and orthographic variations). Table 4.2 shows a breakdown of the frequencies and number of types of each of the three attested senses of *geil* in the *NottDeuYTSch* corpus.

Table 4.2: Frequencies of the three attested senses of *geil* in the *NottDeuYTSch* corpus

| **Sense of 'geil'** | **Frequency** | **Instances per million words** | **Types** |
|---|---|---|---|
| 'sexually aroused' | 196 | 6.02 | 39 |
| 'having a strong desire (towards something)' | 2,261 | 9.46 | 161 |
| 'cool' | 146,645 | 4,505.30 | 1,827 |
| **Total** | **149,102** | **4,580.78** | **2,027** |

Compared to other corpora since 1980 (i.e. since the recognition of the use of

*geil* to mean 'cool'), the frequency of *geil* in the *NottDeuYTSch* corpus is roughly 1,000 times higher than that in either the *DeReKo* (3.55 ipm since 1980) or the *DWDS Kernkorpus* (6 ipm since 1980). The high frequency and productiveness of *geil*, as demonstrated by the data presented in Tables 4.1 and 4.2, make it a prime subject for further linguistic and longitudinal study of youth language. To address the productiveness of *geil*, Section 4.3 presents in-depth statistical analyses of compounding and derivation with *geil* in the *NottDeuYTSch* corpus, while Section 4.4 investigates longitudinal changes in the usage of *geil*.

## 4.3 Innovative compounding and derivation with *geil*

This section presents two quantitative analyses of constructions containing *geil* in the *NottDeuYTSch* corpus to investigate the lexical and morphological developments within digital youth language, demonstrating how the different senses of *geil* appear in different morphological constructions. I first analyse *geil* as an affixoid, particularly as a suffixoid in constructions where *geil* has Sense 3, i.e. 'desperate for something'. The second analysis examines *geil* as the base lexeme in constructions with affixes and affixoids, especially in combination with intensifying prefixes and prefixoids. The second analysis also models the use of multiple intensifying prefixes and prefixoids, which, I argue, is characteristic of German youth language. Before presenting these analyses, I discuss the validity of the term 'affixoid' to describe certain morphemes in constructions containing *geil* observed in the *NottDeuYTSch* corpus.

### 4.3.1 Using the term 'affixoid'

The grammatical concept of the affixoid (also referred to as a 'half-affix,' Elsen 2009: 316) is a "highly controversial topic in German linguistics" (Ascoop & Leuschner 2005: 1), as it highlights the unclear distinction "between the processes of compounding and derivation" (Fagan 2009: 99). Affixoids, such as *Riesen-* in the construction *Riesenparty*, are identical to free lexemes, which

convey meaning on their own (i.e. 'giant'), and so, in combination with other free lexemes, would appear to yield a compound. However, affixoids can have a derivative function, like that of an affix: in the example above, *Riesenparty* actually means a "very big" party and not a "party with giants" (Elsen 2009: 319).

Using examples from the *NottDeuYTSch* corpus, the lexemes *affengeil* and *ungeil* may serve to illustrate the differences between affixoids and affixes. The word *affengeil* (lit. 'monkey cool'), is formed from *der Affe-* ('monkey') + *-n-* (interfix or 'linking element,' see Fagan 2009: 99) + *geil*. *Affe(n)* is classified as a prefixoid as it can be used on its own as a free lexeme, whereas the *un-* of *ungeil* is a bound morpheme, meaning it is only productive in combination with other morphemes or lexemes, and therefore is categorised as a prefix. Ascoop & Leuschner (2005: 3–4), Elsen (2009: 317–19), and Leuschner & Wante (2009: 3) all outlined a series of criteria, which must all be fulfilled for a lexical item to be defined as an affixoid:

1. An affixoid is identical to a free lexeme, e.g. *Riese(n)-* and *der Riese* ('giant').
2. An affixoid is etymologically related to, but semantically more abstract than, the corresponding free lexeme, i.e. *die Riesenparty* refers to a large party, rather than a party involving giants.
3. The affixoid does not determine the "basic meaning" of the lexeme (Ascoop & Leuschner 2005: 3), i.e. *Buschwerk* is a kind of bush, not a kind of creation or factory. This is especially noticeable when the second element is an affixoid (i.e. a suffixoid). For constructions containing a prefixoid and regular compounding, the element containing meaning is expected to be the final element.
4. The affixoid is "highly productive" (Ascoop & Leuschner 2005: 4) and can combine with many other root words, compounds, or base lexemes, similar to an affix, e.g. *Riesenbaum* ('large tree') or *Riesenfan* ('massive fan').

Critics of the notion of affixoids as a separate morphological category, e.g. Schmidt (1987) and Donalies (2005), have argued that the concept is not necessary, as the word building units should be classified as either affixes (i.e. derivation) or lexemes (i.e. compounding). However, as Elsen (2009: 320) noted, "it is

repeatedly stated that a clear demarcation between composition and derivation is not possible" (see also Fagan 2009: 99).[3] For example, as outlined in Ascoop & Leuschner (2005: 3), it is not clear cut whether the construction *Buschwerk* ('undergrowth') is derived from the noun *Busch* ('bush') and the suffix *-werk* ('denotes the totality/large size/extensiveness of something') or whether *Buschwerk* is a [NOUN] + [NOUN] nominal compound (see Example 101a in Fagan 2009: 101).

A possible argument against the validity of the term 'affixoid' to describe the morphological processes involving *geil* arises from Ascoop & Leuschner (2005: 3) and Elsen (2009: 317), who both claimed that prefixoids (as opposed to prefixes) often have "strengthening, intensifying, evaluative" functions, e.g. *Riesenparty* ('a huge party'), while suffixoids (as opposed to suffixes) often have "collective" or "privative" functions, e.g. *Buschwerk* ('undergrowth') or *asbestfrei* ('free of asbestos'). While prefixoids in constructions with *geil*, such as *affengeil*, are indeed intensifying, constructions in the *NottDeuYTSch* corpus where *geil* is a suffixoid, e.g. *likegeil* ('desperate for likes') or *klickgeil*[4] ('desperate for views/clicks') are evaluative rather than collective or privative. However, the functional differences between prefixoids and suffixoids identified in Ascoop & Leuschner (2005) and Elsen (2009) are presented as tendencies, rather than as defining criteria for affixoids. Indeed, the suffixoid *-papst* in *Literaturepapst* ('leading authority in literary criticism'), given in Ascoop & Leuschner (2005: 3), has a strengthening function, denoting authority. Overall, the uncertainty regarding the categorisation of these constructions as either composition or derivation supports the use of separate terminology for this grey area. I therefore use the terms affixoid, as well as prefixoid and suffixoid in this chapter to describe the relevant word-building processes involving *geil* in Sections 4.3.2 and 4.3.3.

[3] A fuller analysis of the arguments against the term 'affixoid' can be found in Elsen (2009).

[4] An important part of *YouTube* content creation, especially in youth-oriented Social Media Entertainment is appearing 'authentic', and (supposedly) creating content for the fun of it, rather than just to earn money (see Burgess & Green 2018: 33-34). If a *YouTube* channel produces content that is interpreted by viewers as being intended primarily to boost their metrics, often referred to as "clickbait" (Zannettou et al. 2018), they can expect a backlash in the comments section, e.g. *du bist einfach nur klickgeil* ('you are simply desperate for views').

### 4.3.2 *Geil* used as an affixoid

There are no attested instances of the use of *geil* as a prefixoid, either reported in previous scholarship, or found in the *NottDeuYTSch* corpus. Mroczynski (2018: 331) claimed that *geilomat* (lit. 'cool-omat') was an example of *geil-* as a prefixoid, but, in this function, it does not fulfil all the criteria of an affixoid listed in Section 4.3, specifically, Criterion 4, that the lexeme created is a combination of the affixoid with an existing lexeme. In fact, *-omat* is not a free lexeme, but functions as a suffix, whereas *geil* here functions as the base lexeme. The use of *-geil* as a suffixoid (e.g. in constructions such as *likegeil* and *klickgeil*, where *like* and *klick* are the base lexemes), however, is productive, combining with 46 different base lexemes in total (i.e. types) and appearing 2,261 times (69.5 instances per million, i.e. tokens) in the corpus.[5] These constructions are more frequent than the combination of intensifying prefixes and prefixoids with *geil* (e.g. *megageil* and *supergeil*, where *geil* is the base lexeme, examined below in Section 4.3.3). This is surprising, given the strong association of intensified forms of *geil*, such as *megageil*, with German youth language (e.g. Androutsopoulos 1998: 103-110). The relatively high frequency and variety of constructions with *geil* as a suffixoid found in the corpus has not been identified in previous scholarship.

The suffixoid *geil* is used in Sense 3 of 'having a strong desire (towards something) or desperate (for something)' in the corpus, most often in a clearly negative context, e.g. *likegeil* ('greedy/desperate for likes'), *geldgeil* ('money-grubbing'), and *klickgeil* ('greedy/desperate for clicks/views'), as shown in the use of *klickgeil* in Example 4.18 below. Specifically sex-related constructions, e.g. *sexgeil* and *cockgeil*, could be interpreted as carrying both Sense 3 ('desperate for something'), and the older Sense 1 ('sexually aroused'). However, I would argue that *-geil*, in these examples carries Sense 3 ('desperate for something'), as the sexual connotations come from the base lexeme, *sex* and *cock*, rather than the suffxoid, *-geil*.

[5]These constructions are not to be confused with the combination of *geil* as a base lexeme with noun-based prefixoids, which often function as intensifiers, e.g. *hammergeil* ('hammer-cool') and *notgeil* (lit. 'emergency-horny', i.e. so sexually aroused that it has become an emergency), and which are examined in Section 4.3.3.

**Example 4.18** (2017)

du bist einfach nur **klickgeil** und hast dur [nur] gedacht oh so kann ich sehr viele Klicks und Geld bekommen [...] das geht echt zu weit😠😠😠[...]😠😠😠 (you are simply **greedy for views** and just thought oh this is how I can get loads of views and money [...] that has really gone too far 😠😠😠[...]😠😠😠)

The base lexemes where *-geil* appears as a suffixoid in the corpus tend to be nouns, although some, such as *like-* and *klick-*, could alternatively be interpreted as verb stems. These are dealt with in Section 4.3.3, which analyses prefixes and prefixoids attached to *geil*, which functions in those cases as the base lexeme. Table 4.3 lists the ten most frequent base lexemes that occur with *geil* as a suffixoid in the *NottDeuYTSch* corpus, grouping together spelling variations and inflected adjective forms. For example, *likegeil* includes variations such as *leikgeilen*, which is both an example of non-standard spelling and inflected for the following noun, appearing in the phrase, *ihr leikgeilen Hurensöne* ('you like-hungry sons of bitches'). Only the top ten of the 40 types of base lexeme appear four or more times, and 50% of the base lexemes appear only once, e.g. *schminkgeil* ('desperate for [putting on] make-up'). Seven of the ten most frequent base lexemes in Table 4.3 are conceptually related to the *YouTube* industry, more specifically direct metrics that determine the success of a *YouTube* channel, e.g. likes (*likegeil*),[6] views (*klickgeil*), and subscribers (*abogeil*), as well as trappings of success, e.g. recognition (*aufmerksamkeitsgeil*, 'desperate for attention') and money (*geldgeil*).[7]

[6] *YouTube* officially uses the terminology 'like' and 'dislike' to refer to ratings received by a video or comment (YouTube 2020), having switched from a 5-star Likert scale system in March 2010. For 'like' and 'dislike', 'thumbs up' and 'thumbs down' icons are used, which explains the occurrence of *daumenhochgeil* ('desperate for thumbs up') twice in the *NottDeuYTSch* corpus.

[7] Roughly 80% of a *YouTuber*'s revenue is earned through advertisements that play before and during videos (run by Google 'Adsense') (*YouTube Premium Is Broken...* 2021: 5:37), and each 1,000 times an advert appears to a viewer (an 'impression'), the *YouTube* channel will receive between €0.20 and €4.00, depending on the cost of the advert.

Table 4.3: The ten most frequent base lexemes with *geil* as a suffixoid in the *NottDeuYTSch* corpus

| Base lexeme | Raw frequency | Instances per million words |
|---|---|---|
| like- | 694 | 21.300 |
| klick- | 651 | 20.000 |
| geld- | 635 | 19.500 |
| abo- | 97 | 2.980 |
| fame- | 89 | 2.730 |
| aufmerksamkeits- | 33 | 1.010 |
| sex- | 8 | 0.246 |
| macht- | 7 | 0.215 |
| konsum- | 6 | 0.184 |
| kamera- | 4 | 0.123 |

### 4.3.3 *Geil* as a base lexeme

There are around 176 different lexemes with *geil* as the base lexeme in the *NottDeuYTSch* corpus (300 if one treats spelled variations, e.g. *geilomatik* and *geilomatic* separately), making this kind of construction the most productive of all compounds in the corpus.[8] Sections 4.3.3.1 to 4.3.3.3 analyse the different

[8]The 176 constructions with *geil* as the base lexeme in the *NottDeuYTSch* corpus are as follows: *affengeil, affenhammergeil, affenhyperdynamitgeil, affenmegageil, affentittengeil, affentittenmegagigantischgeil, alleallegeil, allegeil, apegeil, arschgeil, arschtittengeil, bombegeil, bombevernichtendmegaaffenhammergeil, dauergeil, daumenhochgeil, donnergeil, doppelgeil, dünngeil, echtgeil, einhorngeil, einhornglitzesupergeil, einhornkrassgeil, einhornmegapupsgeil, einhornpupsglitzestickerdonutgeil, endgeil, exorbitantoberaffentittengeil, extrageil, extremgeil, fettgeil, firegeil, fuckingcreepygeil, gagageil, geilheit, geili, geilichkeit, geilig, geilinator, geilo, geilol, geiloman , geilomart , geilomaster , geilomat, geilomatibibi , geilomatic, geilomatico, geilomatikostisch, geilomatrix, geilomeilo , geilometer, geilometra, geilon, geilopower , geiloregalo , geiloritter, geilos, geiloszern, geilotronisch, geilti, geilös, gewinnspielgeil, hammeraffentittengeil, hammergeil, hammerhartgeil, hammermegaaffengeil, hammermegaaffentittengeil, hammermegaobergeil, hammeroberaffentittengeil, hammerobergeil, hammersupergeil, hartgeil, hasengeil, heftiggeil, hochgeil, hodensackgeil, hyperhammergeil, jageil, klargeil, kniegeil, krassgeil, lassengeil, machinegeil, malmgeil, megaabgefucktobermonsterarschfickgeil, megaaffengeil, megaaffentittengeil, megaeinhorngeil, megageil, megahammeraffengeil, megahammeraffentittengeil, megahammerbombesupernakedgeil, megahammergeil, megahammersuperdupergeil, megaheftiggeil, megakrassgeil, megaoberaffengeil, megaobergigaaffengeil, megasuperbombeaffengeil, megasuperdupercoolunglaublichtollgeil, megasuperultrageil, menschgeil, mistgeburtgeil, monstergeil, monstertittengeil, mordgeil, mordhammergeil, notgeil, nougatschnittegeil, nukleargeil, nutellaeinhorngeil, oberaffengeil, oberaffentittengeil, obergalaktischeinhornkakageil, obergeil, oberhammeraffengeil, oberhammeraffentittengeil, oberhammergeil, oberhammermegaaffentittengeil, oberhammermegageil, oberhammertittengeil, obermegaaffengeil, obermegageil, obermegahammergeil, obermegasuperhammergeil, oberschnittegeil, perversgeil, popogeil, pornogeil, richtiggeil, riesengeil, satzzeichenrudelgeil, saugeil, saumegageil, scheißegeil, schnittegeil, schwabbeltittengeil, schwarzershakergeil, schweingeil, schwulgeil, semigeil, sinnlosgeil, sogeil, superaffenarschgeil, superaffengeil, superaffentittengeil, superdupermegaaffentittengeil, supergeil, superhammereinhornglitzepupsgeil, superhammergeil, superhammermegageil, supermegaaffengeil, supermegaaffenhammergeil, super-*

kinds of affixes and affixoids that appear with *geil* in three parts:

1. Section 4.3.3.1 investigates the most productive prefixes and prefixoids used as intensifiers in combination with *geil*, including differences in intensifier use according to the sense of *geil* being used.
2. Section 4.3.3.2 analyses the innovative use of multiple intensifying prefixes and prefixoids within a single lexeme. I present a model to describe the internal grammatical structure of such lexemes with multiple intensifiers.
3. Section 4.3.3.3 examines the most productive suffixes in combination with *geil*, and their functions in comments, in particular the use of intensifying suffixes in comparison with the findings in Mroczynski (2018).

#### 4.3.3.1 *Geil* used with prefixes and prefixoids

There are 150 different combinations of prefixes and prefixoids with *geil* as the base lexeme in the corpus, with the overwhelming majority functioning as intensifiers. Notable exceptions are the negating prefix *un-* in *ungeil* ('uncool', see Example 4.19), and verbal prefixes, such as *auf-* in *aufgeilen* ('get worked up', see Example 4.19). While it would be interesting to investigate verbal prefix constructions more thoroughly,[9] for reasons of space, the following analysis focuses solely on morphological intensification (see Stratton 2020: 186), i.e. intensification through the use of prefixoids and prefixes, such as *super-* in *supergeil* ('super cool', see Example 4.21).

**Example 4.19** (2018)
Der Kuchen sieht einfach SO **ungeil** aus :D ('The cake just looks SO **uncool** :D')

**Example 4.20** (2014)
Ja das sind die kiddies die **sich** daran **aufgeilen** im Video zu sein ('Yes these are the kiddies who **get off** from being in the video')

*megageil*, *superoberhammergeil*, *supersaugeil*, *superwunderbarendgeil*, *technikgeil*, *todgeil*, *totalfreshmegageil*, *ultraaffengeil*, *ultraaffenhammergeil*, *ultrageil*, *unfassbargeil*, *unterleggeil*, *urgeil*, *verdammtgeil*, *vollgeil*, *weltgeil*, *wundergeil*, *zugeil*, *übelgeil*, *übergeil*, *überhammergeil*, *überkrassaffengeil*, and *überübergeil*.

[9] There were 78 (2.4 ipm) instances of verbal prefixes with *geil* in the corpus, covering three types (*aufgeilen*, n = 72; *angeilen*, n = 5; and *begeilen*, n = 1).

**Example 4.21** (2017)
Booooa :D Dieser cliffhänger xD **Supergeil** <3 ('Wowwww :D This cliffhanger xD **Super cool** <3')

The analysis in this section excludes syntactic intensification, where adverbs and particles are used, such as *so* (Example 4.22) and *echt* (Example 4.23), and what I refer to as 'quasi-morphological' intensification, where a prefix or prefixoid is separated from the base lexeme, *geil* by a space, e.g. *über geil*, as shown in Example 4.24

**Example 4.22** (2018)
#️⃣ budebauen #️⃣ Einhorn-Königin #️⃣ **so geil** 🌈🌈🌈🌈🌈😘😍 ('#️⃣ building a den #️⃣ Unicorn queen #️⃣ **so cool** 🌈🌈🌈🌈🌈😘😍')

**Example 4.23** (2018)
[YOUTUBER] du hast ne **echt geile** Stimme!Ohne scherz jetzt! ('[YOUTUBER] you have a **really cool** voice!No joke at all!')

**Example 4.24** (2013)
so ein **über geiles T-shirt** wäre schon nice ('Such a **super cool T-shirt** would be pretty nice')

Quasi-morphological intensification is excluded from further analysis here because the data would have required a large amount of pre-processing for quantitative analysis to differentiate it from syntactic intensification, beyond the scope of what was possible in this thesis.

In examples of morphological intensification, *geil* typically takes the sense of 'cool', e.g. *hammergeil* (n = 740, lit. 'hammer-cool'), although examples where the sense is 'sexually aroused' are also possible, albeit not as frequent, as in the case of *notgeil* (n = 84, lit. 'emergency-horny'). In the *NottDeuYTSch* corpus, 102 different prefixes and prefixoids used with *geil* were identified, which combined to produce 150 different types of lexeme (n = 1890). Of these, 67% (n = 101) were nonce constructions, i.e. they appeared only once, again demonstrating significant lexical diversity and creativity in the expression of enthusiasm in

digital youth language.

Androutsopoulos (1998: 106–8) identified 16 "adadjectival intensifying prefixoids" in his corpus of youth subcultural fanzines, i.e. prefixoids combining with adjectives (*arsch-*, *fuck-*, *flamm-*, *furz-*, *hammer-*, *kack-*, *rappel-*, *ratten-*, *riesen-*, *sack-*, *sau-*, *scheiß-*, *schweine-*, *stink-*, *stock-*, and *turbo-*) and 17 (with some overlaps) "adnominal intensifying prefixoids", i.e. combining with nouns (*Bomben-*, *Dreck-*, *Fick-*, *Hammer-*, *Höllen-*, *Kack-*, *Killer-*, *Kult-*, *Mords-*, *Panne-*, *Riesen-*, *Sahne-*, *Sau-*, *Scheiß-*, *Schrott-*, *Schweine-*, *Spitzen-*). Table 4.4 shows the frequency of each of these prefixoids in combination with *geil* in the *NottDeuYTSch* corpus. Of the 27 combined prefixoids identified by Androutsopoulos, 17 do not appear at all in constructions with *geil* in the *NottDeuYTSch* corpus, such as *furz-*, although all but *flamm-*, *kult-*, and *panne-* do appear in the corpus as intensifiers with other lexemes, e.g. *kackfrech* and *stinkfaul*. Of the 10 prefixoids identified by Androutsopoulos that do appear with *geil*, *hammergeil* is the most frequent (n = 740), and has been mentioned in other investigations of German youth language (e.g. Bahlo & Klein 2017: 176; Vasiljevič 2018: 921). Similarly, *affengeil* (n = 36) has not only been observed in investigations of youth language (e.g. Androutsopoulos 2005: 179; Bahlo & Klein 2017: 161; Bahlo et al. 2019: 60), but also in wider studies of affixoids (e.g. Ascoop & Leuschner 2005: 3; Elsen 2009: 324).

Contrary to the findings of Androutsopoulos (1998) in his study of fanzines, the *NottDeuYTSch* corpus does contain some adnominal intensifying prefixoids that attach to *geil* as an adjective, namely *bomben-*, *fick-*, and *mords-*. Both *bomben-* and *fick-* are productive morphemes in the corpus as prefixoids,[10] but *geil* is the only non-noun lexeme with which they occur in the corpus, further demonstrating the high productiveness of *geil* within German youth language.

Of the six intensifying prefixes investigated by Androutsopoulos (1998: 113–17): *hyper-*, *mega-*, *ober-*, *super-*, *über-*, and *ultra-*, only three are listed as appearing with *geil* (*ober*, *super*, and *ultra*). However, all six are found with *geil* in the

[10]There are around 60 types containing *bomben-* as a prefixoid, e.g. *Bombenvideo* (lit. 'bomb video', i.e. 'great video'), 50 types involving *fick-*, e.g. *Fickdreck* (lit. 'fuck-dirt', i.e. 'a very contemptible person'), although only five with *mords-*, e.g. *mordsmäßig unnötig* (lit. 'murder-wise unnecessary', i.e. 'extremely unnecessary').

Table 4.4: Frequencies of German youth language prefixoids in Androutsopoulos (1998: 106-108) combined with *geil* in the *NottDeuYTSch* corpus

| **Intensifier** | **Raw frequency** | **Instances per million words** |
|---|---|---|
| arsch- | 20 | 0.614 |
| bomben- | 6 | 0.184 |
| fick- | 2 | 0.061 |
| fuck-* | 1 | 0.031 |
| hammer- | 796 | 24.455 |
| mords- | 1 | 0.031 |
| riesen- | 1 | 0.031 |
| sau- | 112 | 3.441 |
| scheiß- | 4 | 0.123 |
| schweine- | 2 | 0.061 |

*Note:* dreck-, flamm-, höllen-, kack-, killer-, kult-, panne-, rappel-, ratten-, sack-, sahne-, schrott-, spitzen-, stink-, stock-, and turbo- do not appear at all as morphological intensifiers in the NottDeuYTSch corpus
* appears in the NottDeuYTSch corpus as 'abgefucktgeil'

*NottDeuYTSch* corpus and, while *hyper-* occurs just once, the remaining five are, in fact, the most frequent prefixes with *geil*. Androutsopoulos (1998) found that the three prefixes, *ober*, *super*, and *ultra*, all functioned as adverbial intensifiers (i.e. syntactic intensification); this use is also common in the *NottDeuYTSch* corpus. For example, *mega* occurs roughly 50,000 times in this function (about 1,500 ipm). What is more, we now have evidence of the lexicalisation of all three prefixes into predicative adjectives and adverbs that can also be intensified, e.g. *echt mega* (Example 4.25), and *voll super* (Example 4.26), and interjections, e.g. *ja ultraaaaaa* (Example 4.27).[11] This is, incidentally, the reverse of a process in Dutch, where syntactic intensifiers such as *so* are undergoing "delexicalization or grammaticalization", i.e. losing lexical meaning and gaining "grammatical functionality" (Hilte, Vandekerckhove, & Daelemans 2019: 301). The two opposite processes underscore the dynamism and constant change at both a lexical and grammatical level in youth languages.

**Example 4.25** (2016)

die Frisur steht dir echt **mega**! ('The hair cut really **mega** suits you!')

[11] A few comments that seemingly used *hyper* as a free-standing morpheme were found, after closer inspection, to reference the 1994 techno song "Hyper Hyper" by Scooter.

**Example 4.26** (2012)
oh ja! das funktioniert voll **super**, danke! wow, ich fühle mich gerade von mir selber verarscht.. o.0 ('Oh yes! That works completely **super**, thanks! wow I feel like I'm taking the mickey out of myself.. o.0')

**Example 4.27** (2017)
Ja **ultraaaaaa** ('Yes **ultraaaaaa**')

Table 4.5 shows the most frequent morphological intensifiers that directly prefix *geil* in the *NottDeuYTSch* corpus. The top ten consists of five prefixes and five prefixoids (prefixoids are starred with an asterisk), with *hammer-* appearing over four times more frequently than the next most frequent, *mega-*. However, a longitudinal examination of the data in the *NottDeuYTSch* corpus suggests that *hammergeil* is no longer such a prominent lexical item in *Jugendsprache*, despite its relatively high overall frequency. Figure 4.3 shows that not only has the overall raw usage of *hammergeil* decreased considerably over the time period covered by the corpus, but the proportion of all *geil* constructions containing *hammer* has also considerably decreased. The notable spike in frequency in 2012 was due to commenters aping the frequent use of the term in a particular video by one *YouTuber*. The overall decrease in the use of *hammergeil* between 2009-2018 is strongly correlated with the decrease in the general use of *geil* in the sense of 'cool', to which I return in more detail in Section 4.4.

Table 4.5: The ten most frequent intensifying prefixes and prefixoids with *geil* in the *NottDeuYTSch* corpus

| Intensifier | Raw frequency | Instances per million words |
|---|---|---|
| hammergeil* | 740 | 22.70 |
| megageil | 173 | 5.31 |
| supergeil | 172 | 5.28 |
| saugeil* | 110 | 3.38 |
| endgeil* | 110 | 3.38 |
| notgeil* | 84 | 2.58 |
| obergeil | 59 | 1.81 |
| allergeil | 54 | 1.66 |
| übergeil | 46 | 1.41 |
| affengeil* | 36 | 1.11 |

* Categorised as a prefixoid

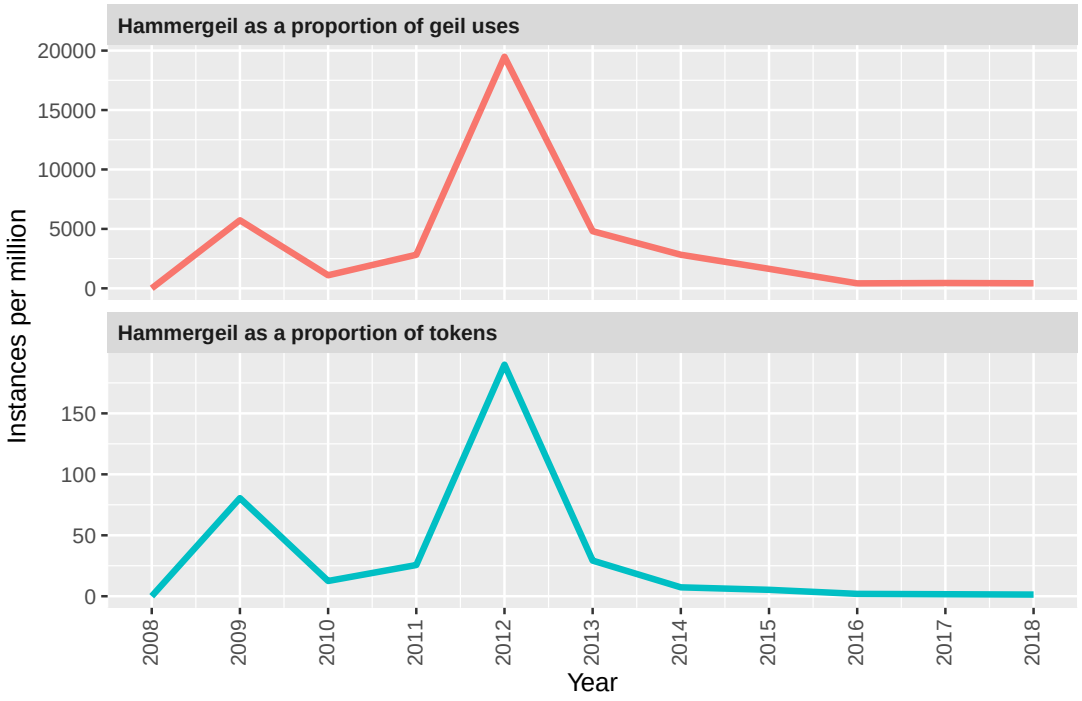



Figure 4.3: Changes in frequency of *hammergeil* over time in the *NottDeuYTSch* corpus

#### 4.3.3.2 Multiple intensifying prefixes and prefixoids with *geil*

Despite the relative decrease in frequency of *hammergeil* over the timeframe of the corpus, *hammer-* has been a very productive prefixoid and was often found in combinations with multiple other morphological intensifiers, such as *ober-*

*MEGAsuperHammerGeilo* (lit. 'upper-mega-super-hammer-cool-o') in Example 4.28. While 93.1% (n = 1,759, 209 types) of intensified *geil* lexemes contained just one morphological intensifier, there were 130 constructions containing a combination of between two and eight intensifiers comprising 97 types.

> **Example 4.28** (2015)
> Hey [YOUTUBER] Es wär **oberMEGAsuperHammerGeilo** wenn ich dabei sein könnte 😁 Ich währe so froh wenn ich beim PVP Battle dabei sein könnt ('Hey [YOUTUBER] It would be **"above-MEGA-super-hammer-cool-o"** if I could be there 😁 I would be so happy if I could take part in the PVP [player vs player] battle')

Table 4.6 shows the ten most frequent morphological intensifiers that appear in constructions of multiple morphological intensifiers with *geil*, the frequency of each intensifier in constructions with *geil* in Column 2 (independent of the number of intensifiers in the word), and the proportion of those constructions out of the total number of intensifying constructions with *geil* as the base lexeme in Column 3. The table also shows the frequency with which an intensifier has appeared in *geil* constructions containing multiple intensifiers (Column 4) and the proportion of constructions containing multiple intensifiers where that intensifier appears (Column 5). For example, constructions containing *hammer-* occur 796 times in total, either as a single intensifier (e.g. *hammergeil*, which occurs 740 times, as per Table 4.5 above) or in any combination of intensifiers (n = 56), e.g. *oberhammergeil* (n = 25, 'upper-hammer-cool'). *Hammer-* occurs in 42.1% of all constructions with *geil* with any number of morphological intensifiers, but only 7% (56 of 796) of constructions containing *hammer-* are constructions with multiple intensifiers. *Hammer-* is, then, mostly used as a standalone intensifier and occurs relatively infrequently in constructions with multiple intensifiers. In contrast, *mega* (24.5%, 56 of 229) and *ober* (47.3%, 53 of 112), for example, occur in proportionally far more combinations with other prefixoids, e.g. *ober-affengeil* (lit. 'above-monkey-cool', n = 6) or *superhammermegageil* (lit. 'super-hammer-mega-cool', n = 1), with *ober* appearing at roughly the same frequency in combinations with other intensifiers and as a standalone intensifier.

Table 4.6: Ten most frequent morphological intensifiers in *geil* constructions containing multiple intensifiers in the *NottDeuYTSch* corpus

| **Intensifier** | **Total frequency (all constructions)** | **Total frequency/all constructions (%)** | **Multiple intensifier frequency** | **Multiple intensifier frequency/total frequency (%)** |
|---|---|---|---|---|
| hammer- | 796 | 42.1 | 56 | 7.0 |
| mega- | 229 | 12.1 | 56 | 24.5 |
| super- | 199 | 10.5 | 27 | 13.6 |
| sau- | 112 | 5.9 | 2 | 1.8 |
| ober- | 112 | 5.9 | 53 | 47.3 |
| end- | 111 | 5.9 | 1 | 0.9 |
| affen- | 93 | 4.9 | 57 | 61.3 |
| not- | 84 | 4.4 | 0 | 0.0 |
| alle- | 56 | 3.0 | 2 | 3.6 |
| über- | 50 | 2.6 | 4 | 8.0 |

Figure 4.4 shows the distribution of the number of morphological intensifiers used in constructions with *geil* as the base lexeme. For example, there are six unique constructions with five prefixoids each in the corpus, such as *oberhammermegaaffentittengeil* ('above-hammer-mega-monkey-tits-cool', n = 3, each time written by a different user) and *superhammereinhornglitzerpupsgeil* ('super-hammer-unicorn-glitter-fart-cool', n = 1).

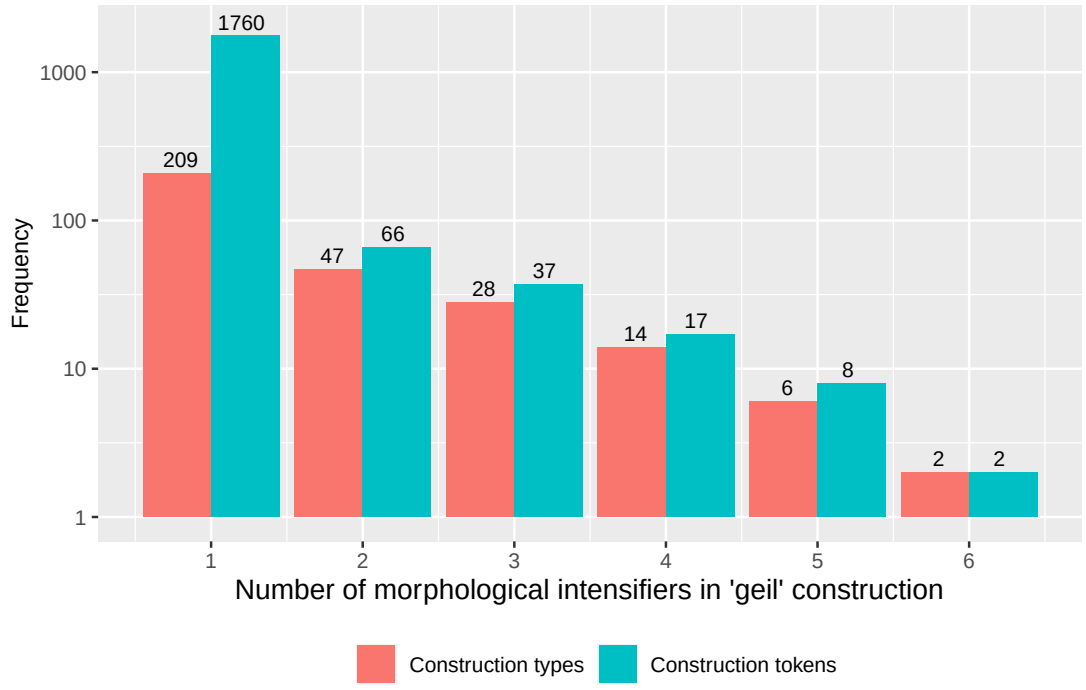



Figure 4.4: Distribution of morphological intensifiers in *geil* constructions

The wealth of data in the *NottDeuYTSch* corpus has made possible additional analyses of the complex constructions containing multiple morphological intensifiers used with *geil* in *Jugendsprache* that have previously not been studied in such depth. In this case, a combination of n-gram[12] and collocation analyses reveals that the order of intensifiers in constructions containing multiple morphological intensifiers is to some extent predictable, with more grammaticalized items occurring earlier, and more lexicalised items later. These complex constructions tend to appear in five classes in the following order:

1. Prefixes that are also used in non-intensifying word formation, such as

[12] An n-gram analysis examines sequences of a given number of items, i.e. a bigram refers to two items, and trigram three items. In the analysis in this section, the items investigated were intensifying prefixes and prefixoids.

*ober* and *über*. These appear as the initial intensifier in 96.2% of cases, e.g. *überkrassaffengeil*.

2. Prefixoids used for intensification, which have been lexicalised as inflectionless adjectives (see Androutsopoulos 1998: 321-323) or adverbs used in syntactic intensification, such as *super* and *mega* (e.g. *supermegageil*).
3. Adjectives, adverbs, and verb participles derived into intensifiers (often conversion/zero-derivation Fagan 2009: 96-97), such as *fresh* and *abgefuckt* (e.g *totalfreshgeil*, *megaabgefucktgeil*).
4. Prefixoids derived from nouns, e.g. *Hammer* and *Affe(n)* (*Hammergeil*).
5. Verb stems, e.g. *fick* (*obermonsterarschfickgeil*).

**Example 4.29** (2015)
ZERSTÖRUNG JAAAAAAA! **Überkrassaffengeil**! ('DESTRUCTION YEAAAAAAH! **over-wicked-monkey-cool**!')

In Example 4.29, *überkrassaffengeil* contains three intensifiers; *über-* (a prepositional prefix, Class 1), *krass-* (adjective, Class 3), and *affen-* (noun-based prefixoids, Class 4), which follow the order above. However, there are several caveats to this order. First, it is not necessary for all kinds of intensifiers to be used in a construction; so an intensifier from later in the order may occur as the first intensifier if no intensifiers from earlier in the order are used, e.g. in Example 4.30, *supersaugeil* contains two prefixoids: an adjectival intensifier (Class 2) followed by a noun-based prefixoid (Class 4), and in Example 4.31, *oberhammeraffentittengeil* has four prefixoids: a standard prefix (Class 1) followed by three noun-based prefixoids (Class 4).

**Example 4.30** (2016)
Sowas von ein **SuperSauGeiles** Video ich lach seit 4 Minuten durchgängig Mach so was unbedingt öfter. selbst meine oma(!) hat gelacht, nachdem ich ihr auch das original gezeigt habe!!!! 👍 ('Such an example of a **Super-Sow-Cool** video I have been laughing straight for 4 minutes You have to do stuff like this more often. Even my grandma(!) laughed after I showed her the original!!!! 👍')

**Example 4.31** (2014)

Oha - Als ich die Naked 3 gesehen hab… Ich glaube mein Herz ist stehen geblieben! Eine der Naked Paletten zu haben, wäre einfach nur – #**oberhammeraffentittengeil**– !! ('Wow - when I saw the Naked 3… I think my heart stopped! Having one of the Naked palettes would simply be –#**upper-hammer-monkey-tits-cool**– !!')

Second, for more complex constructions, the intensifiers tend to subdivide into groups of multiple prefixes, which follow the order identified above. For example, *oberhammermegaaffentittengeil* (lit. 'above-hammer-mega-monkey-tits-cool') can be broken down into two sets of prefixoids, *oberhammer* and *megaaffentitten*, each of which conforms to the order above.[13] Thirdly, sometimes a pragmatic factor overrides the otherwise grammatical order of the prefixoids, where the initial intensifier is selected because it is most relevant to the video or parent comment. This is demonstrated in Example 4.32 with the use of *einhornmegapupsgeil* (lit. 'unicorn-mega-fart-cool'). According to the stages of intensifiers, *megaeinhornpupsgeil* could be expected, however, the comment has been written underneath a video uploaded by *Bonnytrash*, who heavily uses unicorn imagery in her videos (and merchandise). The use of *einhorn-* (along with the 'Unicorn' emoji) is therefore most relevant to the video, as it can be interpreted as a demonstration of affinity or in-group status with the *YouTube* channel, hence its position at the start of the construction.[14]

**Example 4.32** (2017)

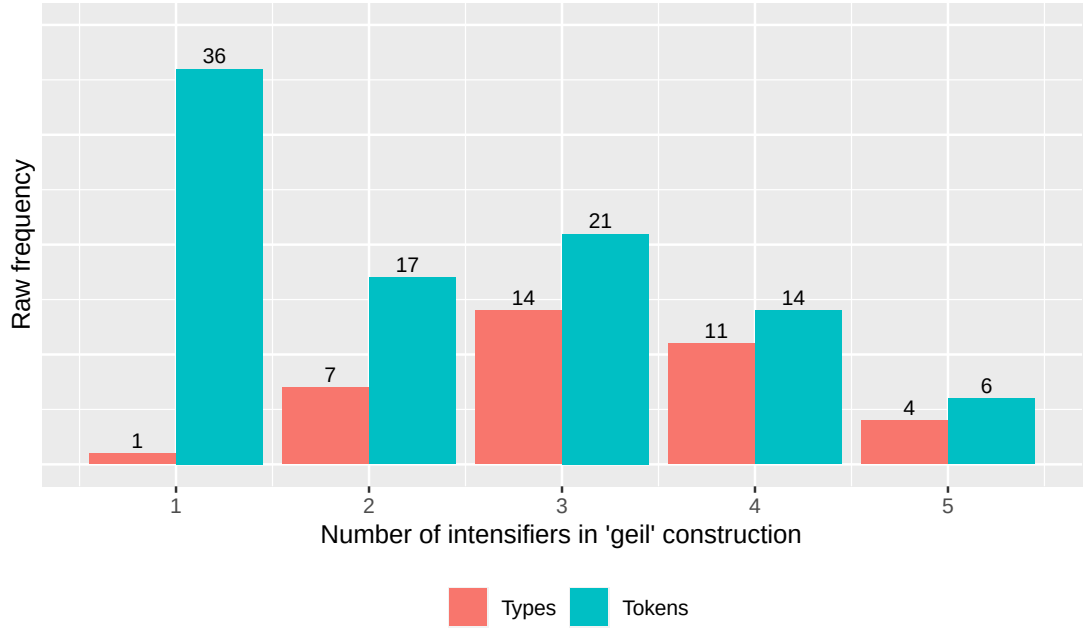
Hey Bonny🦄 Ich bin ein riesennnnnn Fan von dir Ich feier deine videos megaaaaaaaaaa Und ich habe dich abonniert glocke eingeschaltet und

[13] A model instance of the order of prefixoids can be observed in the construction with the largest number of prefixoids in the *NottDeuYTSch* corpus, *obermegafuckinaffenarschundweibertittengeil* (lit. 'above-mega-fucking-monkey-arse-and-women's-tits-cool-shit'), which either has four, five, six, or seven intensifiers, depending on whether one counts *affenarschundweibertitten* (lit. 'money-arse-and-women's-tits') as four separate intensifiers, two compound intensifiers (*affenarsch* and *weibertitten*) or one single phrasal intensifier. Here, the construction consists of a prefix (*ober*, 1), followed by a prefixoid (*mega*, 2), then a verb participle (*fuckin*, 3), and then nouns or noun phrases (*affenarschundweibertitten*, 4).

[14] There are many other linguistic features that are used, like the 'Unicorn' emoji to provide information about aspects of a commenter's identity (such as affiliation to a channel), and a more in-depth examination of such use of language and symbols in the *NottDeuYTSch* corpus is found in Chapter 7.

> jedes video geliket Wäre mege **EINHORNMEGAPUPSGEIL** wen du mich mal in deinem video grüsst […] ('Hey Bonny 🦄 I am a massive fan of yours I enjoy your videos soooooooo much And I have subscribed activated the notification bell and liked every video It would be mega **unicorn-mega-fart-cool** if you could give me a shout out in your video […]')

Even within the 33 million word *NottDeuYTSch* corpus, the use of multiple intensifiers with *geil* is a relatively marginal phenomenon. An analysis of all such compounds with multiple intensifying prefixes and prefixoids remains a desideratum for future research, for example, to expand the testing of patterns or trends within the grammatical order of the intensification model using constructions not restricted to *geil*. However, it was possible to identify some possible emergent trends for the noun-based prefixoid, *affe(n)-*. It is the intensifier most used in combination with other prefixes or prefixoids (n = 57), despite only appearing in 4.9% of all *geil* constructions, and thus appears 61.3% of the time in combination with multiple intensifiers, rather than just on its own (n = 34, 38.7%). Figure 4.5 shows the number of types (in red) and the total number of occurrences (in blue) for every construction containing both *affen-* and *geil* based on the number of intensifiers in the construction.




Note: The construction, 'obermegafuckinaffenarschundweibertittengeil', has been classified as having five intensif
for the purposes of this graph

Figure 4.5: Distribution of number of prefixoids with *affen* in *geil* constructions in the *NottDeuYTSch* corpus

The model of the order of intensifiers in a construction predicts that *affen-*, as a noun-based intensifying prefixoid, is likely to come at the end of any sequence of morphological intensifiers. In fact, however, *affen-* was in the final position, directly before *geil*, in only 47% (n = 27) of cases. A collocation analysis of *affen-* revealed that *affen-* is immediately followed by another noun-based prefixoid, *titten-* ('tits; breasts'), in a further 47% (n = 27) of cases, such as in Examples 4.33 and 4.34, taken from different channels, written in different years.

**Example 4.33** (2014)
[..] WENN ES DICH NICHT GÄBE MIT DEINEN **OBERAFFENTITTENGEILEN** VIDEOS WÄRE ICH ECHT AM BODEN [...] ('[...] IF YOU DIDN'T EXIST WITH YOUR **UPPER-MONKEY-TITS-COOL** VIDEOS, I WOULD BE COMPLETELY GUTTED [...]')

**Example 4.34** (2017)
**OberHammerAffenTittenGeiles**Viedeo :) Weiter so **Over-hammer-monkey-tits-cool**-video :) Keep going

The relatively frequent co-occurrence of *affen-* with *titten-* explains why there are more constructions containing *affen* with three intensifiers than two, as in Figure 4.5. In fact, *affen+titten* is the most frequent pair of consecutive intensifiers in the *NottDeuYTSch* corpus, appearing 28 times (closely followed by *ober+hammer* with 25) and is also part of the most frequent triplet of consecutive intensifiers (*hammer+affen+titten*, n = 9), as shown in Tables 4.7 and 4.8.

Table 4.7: Most frequent pairs of consecutive prefixoids in *geil* constructions in the *NottDeuYTSch* corpus

| **Bigram** | **Raw frequency** | **Instances per million words** |
|---|---|---|
| affen+titten | 28 | 0.860 |
| ober+hammer | 25 | 0.768 |
| ober+affen | 15 | 0.461 |
| mega+affen | 13 | 0.399 |
| hammer+affen | 12 | 0.369 |
| hammer+mega | 12 | 0.369 |
| mega+hammer | 10 | 0.307 |
| super+mega | 8 | 0.246 |
| ober+mega | 7 | 0.215 |
| mega+ober | 5 | 0.154 |

*Note:*
Plus signs added between morphological intensifiers for readability

Table 4.8: Most frequent triplets of consecutive prefixoids in *geil* constructions in the *NottDeuYTSch* corpus

| **Trigram** | **Raw frequency** | **Instances per million words** |
|---|---|---|
| hammer+affen+titten | 9 | 0.277 |
| ober+hammer+affen | 8 | 0.246 |
| mega+affen+titten | 7 | 0.215 |
| ober+hammer+mega | 7 | 0.215 |
| hammer+mega+affen | 5 | 0.154 |
| ober+affen+titten | 5 | 0.154 |
| mega+ober+affen | 3 | 0.092 |
| mega+affen+hammer | 2 | 0.061 |
| mega+hammer+affen | 2 | 0.061 |
| ober+mega+hammer | 2 | 0.061 |

*Note:*
Plus signs added between morphological intensifiers for readability

The use of morphological intensifiers observed in the *NottDeuYTSch* corpus

demonstrates the combination of several previously noted features of German youth language and online communication. It combines innovative compounding and derivation (Bahlo et al. 2019: 59) and often taboo subjects (Bahlo 2012) from youth language, with salient features of digital writing, such as the repetition of graphemes (Dürscheid, Wagner, & Brommer 2010: 134), as demonstrated in Example 4.35, to create idiosyncratic lexical items. These can sometimes reach absurd lengths, far beyond what might be considered necessary to convey the emotion (the functions of such repetition are investigated further in Chapter 7), as evidenced by the complex constructions in Table 4.9. However, the variety and frequency of the use of prefixoids for intensification and for the creation of new compounds illustrate the inventive ways that young people have found to express themselves in an online space.

**Example 4.35** (2016)
Kllllllllllliiiiiiiicccckkkkkkgggggeeeeiiiiiillll … ('Cllllllllllliiiiiiiicccckkkkkkbbbbbaaaaiiiiittttt …')

Table 4.9: A selection of constructions with a large number of prefixoids preceding *geil* in the *NottDeuYTSch* corpus

| **Construction** | **Prefixoids** | **Raw frequency** | **Instances per million words** |
|---|---|---|---|
| affen+titten+mega+gigantomatisch+geil | 4 | 1 | 0.031 |
| ober+hammer+affen+titten+geil | 4 | 4 | 0.123 |
| bomben+vernichtenden+mega+affen+hammer+geil | 5 | 1 | 0.031 |
| einhorn+pupsi+glitzer+sticker+ donut+geil | 5 | 1 | 0.031 |
| mega+hammer+bombig+super+naked+geil | 5 | 1 | 0.031 |
| ober+hammer+mega+affen+titten+geil | 5 | 3 | 0.092 |
| mega+super+duper+cool+unglaublich+toll+geil | 6 | 1 | 0.031 |

*Note:*

Plus signs added between morphological intensifiers for readability

#### 4.3.3.3 *Geil* used with suffixes

As opposed to the frequent and complex nature of prefixoids preceding *geil*, no constructions formed with a suffixoid attached to *geil* are attested in the *NottDeuYTSch* corpus, although there were several creative plays on words (see Bahlo et al. 2019: 104), such as the portmanteau word *geilonachtsmann*, a pun on der *Weihnachtsmann* ('Father Christmas') (n = 10, 0.307 ipm). As for suffixes attached to *geil*, Table 4.10 shows the six most frequent in the corpus (with spelling and thematic variations grouped, e.g. *o* and *oooo*, or *geilomat* and *geilomatico*).[15]

Table 4.10: Most frequent suffixes with *geil* in the *NottDeuYTSch* corpus

| Suffix | Orthographic variations | Raw frequency | Instances per million words |
|---|---|---|---|
| geilo | 66 | 1 737 | 53 |
| geilomat(ic)(o) | 28 | 134 | 4 |
| geilheit | 25 | 107 | 3 |
| geili/y | 13 | 29 | 1 |
| geilkind | 3 | 11 | 0 |
| geilonachtsman | 2 | 10 | 0 |

Suffixes such as *-heit* are used to derive *geil* from an adjective into a noun, often in combination with other compound elements, e.g. *geldgeilheit*, but the most frequent suffix is *-o*, e.g. *geilo* (Example 4.36). The word *geilo* occurs far more frequently than any constructions of prefixes or prefixoids with *geil*, more than twice as often as the most frequent prefixoid combination, *hammergeil* (analysed in Section 4.3.3.1). The *-o* suffix is an example of what Androutsopoulos (1998: 124–26) referred to as "parasitic suffixes", i.e. examples of creative language use that have "no semantic functional value" and "are merely 'attached' to a morphologically unchanged lexeme". This interpretation is perhaps also applicable to *geilomat* and *geilomatico*, although it could be argued that the

[15] In addition to the forms listed in Table 4.10, there were 41 instances of *geila* in the corpus. This is a phonetic spelling of *geiler*, i.e. either inflected as part of a attributive adjective for the gender of the following noun (e.g. "gscheid Geila Track", 'proper cool track'), or inflected for the comparative (e.g. "geila ois ds original", 'cooler than the original'). As seen in the examples, *geila* seems to be frequently used alongside other markers of Austrian German, e.g. *gscheid* (*gescheit*, 'clever', used as an intensifier), and *ois* (*als*, 'than'). However, as *-a* here is used as a suffix of inflection, it is not further examined.

suffixes do indeed both morphologically alter the lexeme and have a semantic functional value. A quantitative analysis of the *NottDeuYTSch* corpus offers tentative evidence that the *-o* suffix may be a contributor to the morphological alteration of *geilo* from an adjective (both predicative and attributive, as in Examples 4.36 and 4.37) to an interjection (Example 4.38).[16] While *geil* also appears without a suffix as an interjection (see Example 4.39), 20% of *geilo* instances are as an interjection (278 out of 1,393 instances) compared to 17% of *geil* instances as an interjection (12,746 out of 75,025 instances), a modest but nevertheless statistically significant difference, $\chi^2$ (1, n = 76,418) = 8.5205, p < 0.01.

**Example 4.36** (2013)
Son T-Shirt ist schon **geilo** :D ('such a T-shirt is pretty **coolio** :D')

**Example 4.37** (2014)
richtig richtig **geilo** viedio meine lieben joonge ('really really **coolio** video my dear maaate')

**Example 4.38** (2017)
**Geilo**!!Zwar erst jetzt entdeckt-aber das wird dieses Jahr als Weihnachtsrundgruß verschickt-perfekt! ('**Coolio**!!Only now just discovered-but this will be sent this year as a mass Christmas greeting-perfect!')

**Example 4.39** (2016)
**geil**!!! aber wie alt wart ihr da? ('**cool**!!! but how old were you then?')

Furthermore, *-o* suffixation can also have a functional semantic value. Mroczynski (2018: 334) claimed that affixes such as *-o* and *-omatico* can be used in compounds with *geil* as a "neue Steigerungsform" ('a new form of comparison'). Specifically, Mroczynski suggested that *geilo* was used as a comparative and *geilomatico* as a superlative to express the degrees of comparison for *geil*. A similar position was taken by Bahlo et al. (2019: 60), who claimed that *megageil* functioned as a "künstlicher Elativ" ('artificial elative/absolute superla-

[16] These interjections can also function as discourse markers. A full discussion of the definition of a discourse marker as used in this study is found in Chapter 6, Section 6.2.

tive'). However, such comparative and superlative forms are not attested in the *NottDeuYTSch* corpus in sentences containing direct comparison, e.g. **Sie ist geilo als er*, nor do such constructions occur in Mroczynski (2018) or Bahlo et al. (2019). Indeed, as demonstrated in Example 4.40, *geilo* can be inflected for the superlative, i.e. *das geiloste*, thus confirming the function of the suffixes as intensifiers rather than as genuine inflections for degree.

**Example 4.40** (2014)
Ich hab alle folgen gesehen und ich muss sagen das mit bergi is nice aber nicht das **geiloste** an freedom aber ich liebe es und like jeden part süchtig durch ('I have watched every episode and I have to say the thing with bergi is nice but the **coolio-est** for freedom but I love it and go through and addictedly give a 'like' to every part')

I argue that in fact, Mrocyznski and Bahlo et al. have actually demonstrated that the more morphologically complex the variation of *geil*, the higher the intensity of emotion, rather than any new grammatical phenomena, i.e. this is a case of morphological intensification.[17]

Morphological intensification involving *geilo* is demonstrated in Example 4.41, where increasingly intensified forms of *geil* are consecutively listed, e.g. *geilo* is less intense than *geilomatico* but more intense than *geil*. Admittedly, *geiler*, the standard comparative form of *geil*, does appear second in the list in Example 4.41, which could lend weight to the suggestion of a new form of comparatives made by Mroczynski (2018), but given the lack of an explicit point of comparison, *geiler* is more likely to function here as an intensifier than *geilo* and *geilomatico* are to function as a comparative and superlative, respectively.

**Example 4.41** (2014)
Geil Geiler Geilo Geilomatico ach [YOUTUBER] ('Cool Cooler Cool-o Cool-omatico ah [YOUTUBER]')

[17]The *-o* suffix can also be used in orthographic intensification if it is repeated, e.g. *geilooooo*, a point to which I return in Chapter 7.

## 4.4 Use of *geil* over time in the *NottDeuYTSch* corpus

Having examined morphological developments of *geil* constructions, this section presents a corpus linguistic investigation of the use of *geil* over time in the *NottDeuYTSch* corpus. I first analyse overall trends and then each of the three senses of the word attested in the corpus individually, before focusing specifically on Sense 4 of *geil*, i.e. 'cool'.

Despite the 40-year history of *geil* as a frequent lexical item in German youth language, the data in Figure 4.6 indicate that *geil* is dropping out of the vernacular of young people. Bahlo et al. (2019: 56) noted that lexical items in youth language can either become a staple over an extended period of time or disappear to the periphery. It seems that *geil* has undergone the first process but is now perhaps undergoing the second: in 2008, it appeared in over 200 tokens per 1,000 in the *NottDeuYTSch* corpus, but its frequency has now dropped to around 30 tokens per 1,000, as shown in the top line of Figure 4.6.

However, the middle line of Figure 4.6 shows that there are some spikes in the raw frequency for *geil*, most noticeably in the Decembers of 2012, 2015, 2016, and 2017. An in-depth investigation of the videos and comments produced in these months revealed that December is a popular time for the production of higher-quality or comment-garnering content for Christmas by *YouTube* channels such as *BibisBeautyPalace* (e.g. https://www.youtube.com/watch?v=z5x0hWmzYv8) and *Julien Bam* (e.g. https://www.youtube.com/watch?v=wANrY7PabWE). The number of comments accordingly increases dramatically (Figure 4.6, bottom), including comments containing *geil*, but the percentage of comments containing *geil* does not consistently show similar increases, apart from in 2012 (Figure 4.6, top).

The usage frequencies of the three attested senses of *geil* (Sense 1 – "sexually aroused", Sense 3 – "desperate for something", and Sense 4 – "cool", outlined in Section 4.1), do not all follow the same pattern, as Figure 4.7 shows. Usage of the most frequent sense of *geil*, 'cool', has experienced a significant drop,

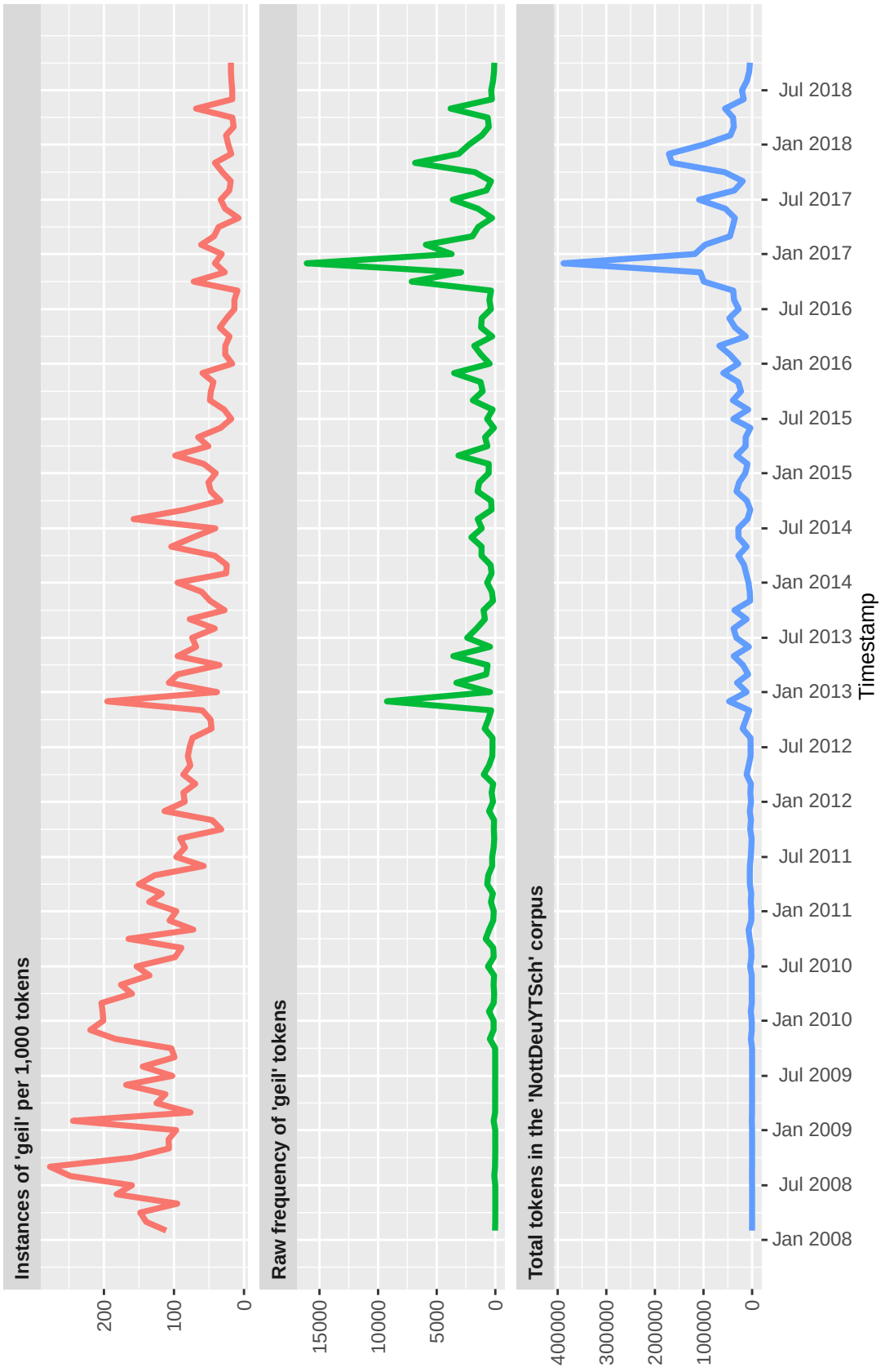



Figure 4.6: Changes in frequency of *geil* over time in the *NottDeuYTSch* corpus

from around 192 tokens per 1,000 in 2008 and 2009 to 29 tokens per 1,000 in 2018. However, the use of *geil* to mean 'sexually aroused' experienced a slight decrease between 2008 and 2010 but then levelled off, and appeared 100-200 times less frequently than the sense of *geil* to mean 'cool'. The use of the sense of *geil* to indicate a strong desire or desperation for something, e.g. *likegeil* ('desperate for likes'), contrary to the other two meanings, has increased over the time period of the corpus. The relative frequency was still far lower than the 'cool' meaning of *geil* throughout the time frame of the corpus, but it surpassed the usage of *geil* with the sense of 'sexually aroused' in 2011. The increase in usage of this last sense of *geil* seems to be connected with the rise of the *YouTube* industry, as examined in Section 4.3.2, since the vast majority of constructions using this sense of *geil* are related to aspects of running a *YouTube* channel, presumably an artefact of the corpus data, collected from *YouTube* comments. In other corpora, the sense of *geil* to indicate desperation is attested in other domains, although infrequently, for example, in politics, *Schwarz-Rot-Geil*, to describe those favouring a grand coalition between the CDU and SPD, or in sport, *torgeil*, meaning desperate for goals (examples taken from the *DeReKo*). If *geil* with Sense 4 ('cool') continues to experience such a drop in frequency of use, the use of *geil* as a suffixoid to indicate desire or desperation looks set to become the most frequent sense in the next decade or so.

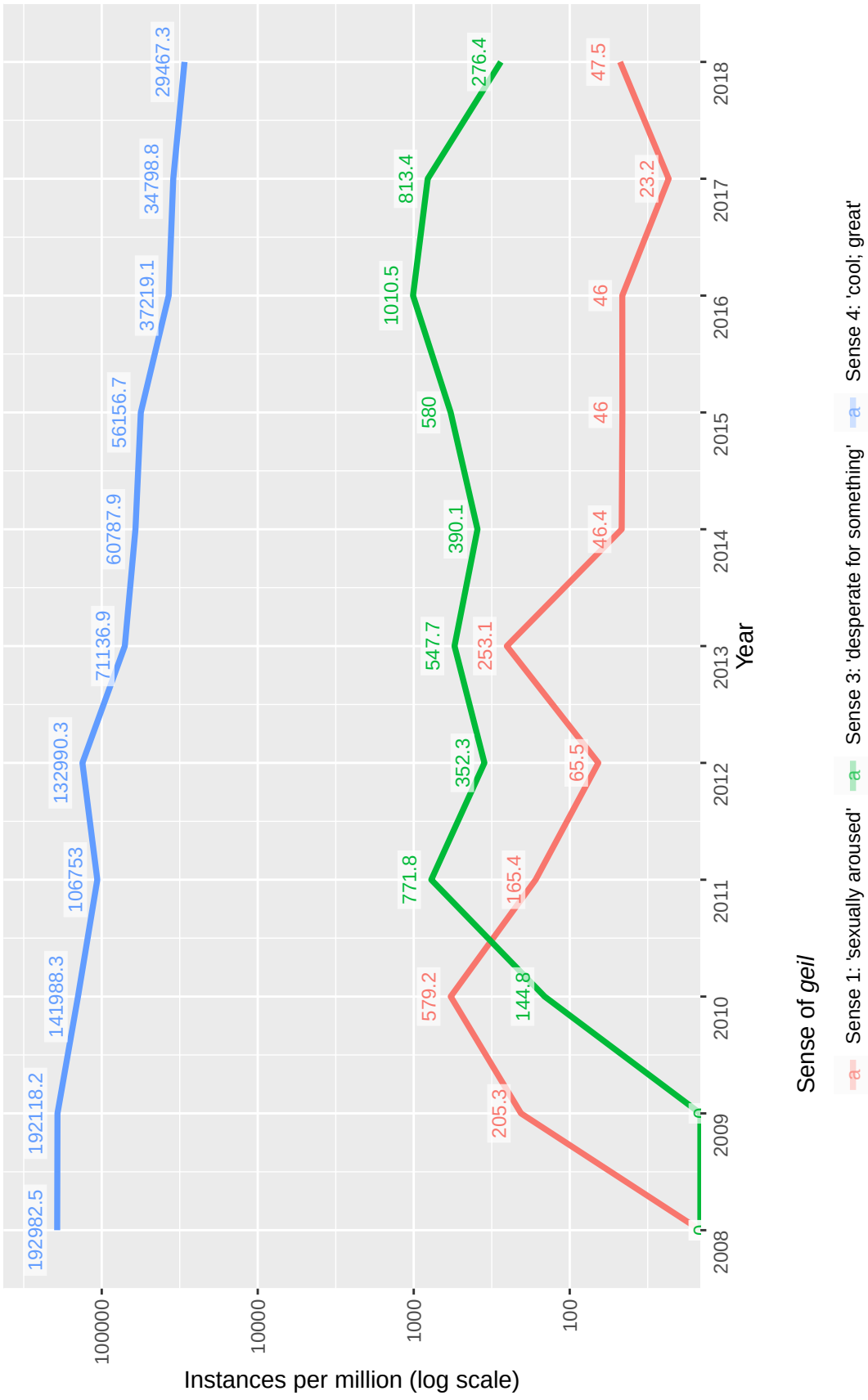



Figure 4.7: Changes in frequency of different meanings of *geil* over time in the *NottDeuYTSch* corpus

### 4.4.1 Alternatives to *geil* in digital youth language

What does the rapid decline in the use of *geil* in Sense 4 to mean 'cool' indicate about potential future expressions of positive emotion of young German-speakers? Evaluative expressions, central to German youth language, remain common in comments in the *NottDeuYTSch* corpus, but other lexical items have seen a considerable increase in usage for this purpose. Section 4.2 presented a list of other expressions of positive emotion identified in previous scholarship. The frequency of these terms in the corpus, as well as of others not examined as adjectives or interjections in previous research (e.g. *heftig* and *mega*) is shown in Table 4.11, which includes three snapshots of the frequency of use (in instances per million) at the beginning, middle, and end of the corpus time range.

Table 4.11: Frequency of positive expressions in the *NottDeuYTSch* corpus listed in alphabetical order (instances per million)

| **Expression** | **2009** | **2013** | **2018** |
|---|---|---|---|
| astrein | 0 | 0 | 0.0 |
| baba | 0 | 94 | 73.0 |
| bombe | 112 | 143 | 41.0 |
| cool | 3 927 | 4 251 | 4 654.0 |
| easy | 32 | 16 | 31.2 |
| episch | 64 | 149 | 34.0 |
| fett | 338 | 198 | 150.0 |
| fresh | 32 | 134 | 88.0 |
| geil | 15 098 | 6 716 | 3 398.0 |
| genial | 1 932 | 316 | 43.0 |
| göttlich | 64 | 11 | 11.0 |
| hammer | 2 447 | 1 144 | 207.0 |
| hart | 338 | 233 | 205.0 |
| heftig | 97 | 88 | 162.0 |
| klasse | 612 | 469 | 314.0 |
| korrekt | 80 | 55 | 61.0 |
| krank | 772 | 977 | 407.0 |
| krass | 499 | 458 | 437.0 |
| mega | 322 | 968 | 2 198.0 |
| nett | 418 | 329 | 171.0 |
| nice | 482 | 465 | 844.0 |
| okay | 418 | 115 | 78.0 |
| scharf | 1 | 24 | 21.0 |
| stark | 64 | 96 | 94.0 |
| super | 917 | 2 122 | 741.0 |
| toll | 1 593 | 2 263 | 770.0 |

Three lexemes have increased in frequency over the course of the *NottDeuYTSch* corpus: *cool*, *nice*, and *mega*, while most have either undergone a constant decrease (e.g. *fett*, *genial*, *klasse*, and *hammer*/*Der Hammer*), or peaked during the period of the corpus and then decreased (e.g. *episch*, *super*, and *toll*). Table 4.12 shows that the top five most frequent positive expressions were different at each of the three census points across the corpus, although *geil*, *cool*, and *toll* have all remained in the top five. However, the frequency of *geil* has roughly halved at each point, and by the final year covered by the corpus, 2018, *cool* had, to judge by the corpus, become the most frequent expression to convey positive emotion among young German speakers.

In fact, *cool* overtook *geil* in 2014, as demonstrated in Figure 4.8, a more detailed chart of the rise of *cool*, *nice*, and *mega*, plotted against the frequency of *geil*. *Cool*, like *geil*, has also previously been described as part of the general slang lexicon of German, rather than a lexical item particular to German youth language (see Androutsopoulos 2005: 175). However, the resurgence of this term, as illustrated by the increase in usage over the time frame of the *NottDeuYTSch* corpus, could indicate that it is being recontextualised as youth language as part of the cycle of lexical renewal and innovation that is characteristic of youth language (see Chapter 2, Section 2.2.2).

Table 4.12: Frequency of positive expressions in 2009 (left), 2013 (centre), and 2018 (right) in the *NottDeuYTSch* corpus

| **Expression (2009)** | **Instances per million** | **Expression (2013)** | **Instances per million** | **Expression (2018)** | **Instances per million** |
|---|---|---|---|---|---|
| geil | 15 098 | geil | 6 716 | cool | 4 654 |
| cool | 3 927 | cool | 4 251 | geil | 3 398 |
| hammer/der Hammer | 2 447 | toll | 2 263 | mega | 2 198 |
| genial | 1 932 | super | 2 122 | nice | 844 |
| toll | 1 593 | hammer/der Hammer | 1 144 | toll | 770 |

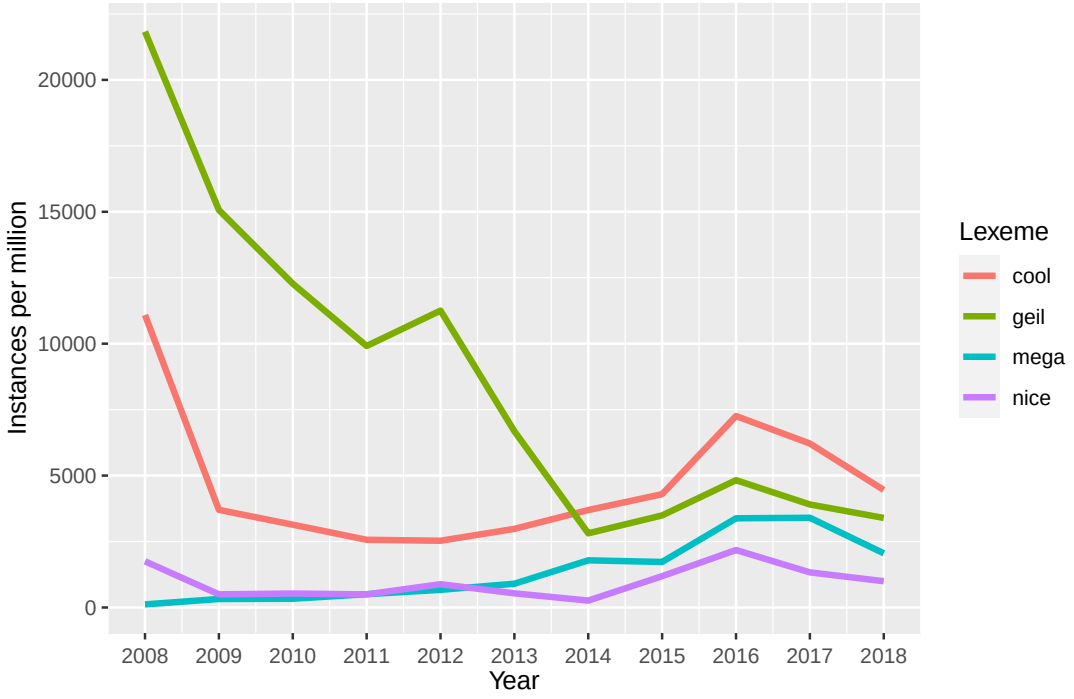



Figure 4.8: Changes in frequency of positive expressions in the *NottDeuYTSch* corpus

From 2015 to 2018, the frequencies of the four lexemes, *cool*, *geil*, *nice*, and *mega*, were very close, evidence that the positive expressions in *Jugendsprache* have been diversifying. From 2016 there is a noticeable downward trend in frequency of all four positive expressions. One potential reason for this is that young people seem to have become less enthusiastic about *YouTube* videos overall, and more likely to express discontent, as is seen with the increase in the use of the negatively valenced constructions with *geil* as a suffixoid, shown in Section 4.3.2 (e.g. *likegeil*). Alternatively, it may be that *geil* and other positive adjectives are being replaced by different or more complex grammatical structures. In the *NottDeuYTSch* corpus, it is noticeable that a number of verb participles and phrases are used to express satisfaction with a video through the description of an action rather than the direct expression of an emotion (statistics from 2018): *abonniert* ('subscribed', 1,048 ipm, n = 3,061), *glockeaktiv/glockeaktiviert/glocke* ('bell active/bell activated/bell', 916 ipm, n = 2,674), and *daumen hoch* ('thumbs up', 427 ipm, n = 1,247).[18]

[18] '*Glockeaktiviert*' refers to the bell icon under a video that is pressed to activate push notifica-

## 4.5 The future of *geil* in digital youth language

The changes in the use of *geil* in its various senses and changes in the preferred positive evaluative expressions in the *NottDeuYTSch* corpus lend support to previous claims in scholarship that German youth language is constantly undergoing lexical innovation and renewal (Bahlo et al. 2019: 56; Androutsopoulos 2005). The data in the *NottDeuYTSch* corpus, with the high and increasing frequency of *cool* and *nice*, overtaking the native German term *geil*, also suggests possible increased influence from the English language. It remains to be seen whether one term will rise to a position similarly dominant to that of *geil* between 1980 and 2010; whether multiple lexemes will be used with a similar frequency; or, indeed, whether there will be a rise in use of words from community languages to express positivity, such as Turkish, e.g. *baba/babo* ('father/boss', derived as an adjective expressing positivity), which peaked in the corpus in 2013 with 94 ipm (n = 241), before experiencing a slight fall to 73 ipm (n = 212). Auer & Dirim (2003), Deppermann (2007), and Wiese & Freywald (2019) have all claimed that community languages are increasingly contributing to youth language in Germany, but such borrowings are, as yet, still infrequent in the comments under mainstream *YouTube* videos captured in the *NottDeuYTSch* corpus. Indeed, given the corpus's deliberate focus on mainstream video comments, some of the more innovative linguistic features of contemporary German youth language used by smaller youth scenes and subcultures are less likely to have been captured in the *NottDeuYTSch* corpus, which might not attract these demographics.

The rise of the use of Sense 3 of *geil* as a suffixoid to express desire or desperation for something requires further investigation with reference to additional data outside of a *YouTube* context. There is also nascent evidence that particular *-geil* compounds that occur in this sense are addressed at *YouTubers* based on their gender identity, with *YouTubers* identifying as men more likely (n = 375) to face accusations of being *geldgeil* ('desperate for money') or *famegeil* ('fame-hungry'), than *YouTubers* identifying as women (n = 43), $\chi^2$ (1, n = 29,797,099)

tions for that *YouTube* channel and '*daumen hoch*' refers to the thumbs up icon that is press to 'like' a video or comment.

= 77.1267, $p < 0.001$.

Despite the fall in relative frequency of *geil* over the time period covered by the *NottDeuYTSch* corpus, *geil* remained in the top 50 most frequent lexemes of the corpus in 2018, and its use amongst young people was far higher than in the everyday language recorded in contemporary reference corpora, such as the *DeReKo* and *DWDS-Kernkorpora*. As the findings in this chapter demonstrate, *geil* has been and remains an extremely productive element of the online language of young German-speakers and a source of creativity. As shown in Section 4.3.2, the use of *geil* as a suffixoid with Sense 3 to mean 'desperate for something', especially in combination with base lexemes, such as *Like* and *Geld*, illustrates innovation to create new lexemes to describe the lived experience of young people, in this case, engagement with the discourses of the *YouTube* industry. Furthermore, as demonstrated in Section 4.3.3, *geil* has been the base lexeme for a wide range of constructions with several different senses, enabling creative morphological intensification through both the use of multiple (often prefixing) intensifiers and the complexity of the affix (often with suffixes, e.g. *geilomat*), which provides an illuminating insight into the linguistic processes in both *Jugendsprache* and digital writing. It remains to be seen whether *geil* will remain as pervasive and hence as iconic for another 40 years.

# Part II

# Syntactic developments in digital youth language: subordinating conjunctions

# 5

# *Weil es kamen viele neue updates* – emergent non-standard word order following subordinating conjunctions

Having analysed selected lexical and morphological developments in digital youth language in the *NottDeuYTSch* corpus in Chapter 4, this chapter and the next turn to syntactic developments, in particular the use of subordinating conjunctions in clauses with non-standard word order. The present chapter first examines what is known from previous research about this phenomenon, and then describes the process of preparing and identifying the data in the *NottDeuYTSch* corpus for analysis of this non-standard syntactic phenomenon. We shall see, first, that the use of non-standard word order in clauses introduced by subordinating conjunctions increased in the corpus over the period of 2008 to 2018. Second, compared to the data in previous scholarship, the corpus provides evidence of additional subordinating conjunctions that introduce this syntactic phenomenon. The theoretical and methodological background, together with the initial findings from the corpus, presented in this chapter,

lay the foundation for more detailed analyses of non-standard word order in clauses introduced by subordinating conjunctions in Chapter 6, which investigates the grammatical processes that could explain the developments outlined in this chapter.

To analyse the changes to the word order of what have traditionally been considered subordinate clauses, I use the "topological model" of German syntax (Fagan 2009: 138-139; Leibniz-Institut für Deutsche Sprache 2017b) to provide a framework to explain the position of elements within a sentence or clause as part of the analyses. The topological model divides a sentence or clause into seven possible fields, although not all fields are necessarily present in any one example. They are, occurring from left to right in a clause, the left margin (*linkes Außenfeld*), first field (*Vorfeld*), left bracket (*linker Satzklammerteil*), midfield (*Mittelfeld*), right bracket (*rechter Satzklammerteil*), final field (*Nachfeld*), and right margin (*rechtes Außenfeld*). Table 5.1 shows three example sentences from Leibniz-Institut für Deutsche Sprache (2017b) arranged according to their classification in the model. The final field and right margin occupy the same column, as their elements can appear in the same position within a sentence, and even overlap (see Leibniz-Institut für Deutsche Sprache 2018b).

Table 5.1: The topological model of German syntax, adapted from Leibniz-Institut für Deutsche Sprache (2017)

| **Left margin** | **First field** | **Left bracket** | **Midfield** | **Right bracket** | **Final field/ right margin** |
|---|---|---|---|---|---|
| Also | das | hat | mich wirklich | überrascht | |
| Den Uwe, | den | treffe | ich häufiger, | | Papa. |
| | Wir | haben | das nicht | gemacht, | wirklich! |

*Translations:*
[1] Well, that really surprised me
[2] Uwe, I bump into him [Uwe] more often, Dad
[3] We didn't do it, honestly!

In standard German, subordinating conjunctions, such as *weil* ('because') and *obwohl* ('although'), require a hypotactic word order, i.e. the finite verb occupies the right bracket position in the clause, often found at the very end of the clause.

An example sentence with hypotactic word order is shown in Table 5.2.

Table 5.2: Topological field diagram for a hypotactic clause

| **Main Clause** | | **Left Bracket** | **Midfield** | **Right Bracket** |
|---|---|---|---|---|
| Ich muss schlafen | , | *weil* | ich müde | *bin* |
| I must sleep | , | *because* | I tired | *am* |
| I must sleep | (,) | because I am tired | | |

Coordinating conjunctions, on the other hand, such as *und* ('and') and *denn* ('because/since/for'), require a paratactic word order, i.e. the finite verb is in the left bracket position, as is the case for main clauses. Again, an example sentence demonstrating paratactic word order is shown in Table 5.3.

Table 5.3: Topological field diagram for a paratactic clause

| **Main Clause** | | **Left Margin** | **First Field** | **Left Bracket** | **Midfield** |
|---|---|---|---|---|---|
| Ich muss schlafen | , | *denn* | ich | *bin* | müde |
| I must sleep | , | *for* | I | *am* | tired |
| I must sleep | (,) | for I am tired | | | |

However, some subordinating conjunctions can also be used with a paratactic clause structure, which although non-standard, is reportedly a characteristic of informal, spoken German (Durrell 2003: 22). This non-standard paratactic structure is illustrated in Table 5.4.

Table 5.4: Topological field diagram for a 'subordinating' conjunction with paratactic word order

| **Main Clause** | | **Left Margin** | **First Field** | **Left Bracket** | **Midfield** |
|---|---|---|---|---|---|
| Ich muss schlafen | , | *weil* | ich | *bin* | müde |
| I must sleep | , | *because* | I | *am* | tired |
| I must sleep | (,) | because I am tired | | | |

The phenomenon of non-standard paratactic clause structure has been variously referred to within scholarship as "Nicht-VL-Sätze mit Nebensatzmarkern" ('Non-verb-last-clauses with subordinating markers'), "V2-Nebensätze" ('verb-second subordinating clauses'), "eingebettete V2-Sätze" ('embedded verb-second clauses'), and "V2-Komplementsätze" ('V2 complement clause') (Freywald 2016: 329). The phenomenon has been shown in previous studies, as

discussed in Chapter 2, Section 2.2.4, to occur primarily with clauses introduced by *weil* and *obwohl*, and usage has been reported as increasing since the 1980s (Günthner 1996: 323), although empirical data are lacking. This chapter uses the data in the *NottDeuYTSch* corpus to quantitatively demonstrate not only that a wider range of conjunctions than just *weil* and *obwohl* are used to introduce paratactic clauses, but that the frequency of parataxis has increased considerably over the time frame of the corpus. Furthermore, the data show that parataxis in clauses introduced by these conjunctions is most frequent in digital youth language compared to other digital writing and standard language. The chapter has two sections:

Section 5.1 provides an in-depth explanation of the methodology used to sample the data in the *NottDeuYTSch* corpus and to prepare it for both qualitative and quantitative analyses of parataxis following subordinating conjunctions. As we shall see, the data required a great deal of cleaning and filtering. This made it necessary to restrict the number of subordinating conjunctions examined as well as to narrow the focus to clauses where the finite verb was among the twenty most common verbs. The section explains the reasoning behind the selections, and outlines the steps taken to remove false positives from the data sample, as well as the ambiguous clauses where it was not possible to categorise the syntax structure as either hypotactic or paratactic.

Section 5.2 examines developments in the use of parataxis following subordinating conjunctions within the *NottDeuYTSch* corpus data. The section demonstrates that subordinating conjunctions that have not been examined in-depth in previous scholarship on parataxis in fact have been used in digital youth language to introduce a paratactic clause structure, e.g. *wenn* ('when/if'). The section then presents a longitudinal analysis of parataxis, showing that the use of this phenomenon has significantly increased over the time period of the corpus, and investigates whether register is the key variable to determine the frequency of parataxis, comparatively analysing the data in the *NottDeuYTSch* corpus with data from three other corpora: the *DWDS-Kernkorpus 21*, the *DWDS-WebXL Korpus*, and the *Dortmunder Chat-korpus*.[1]

[1]See Chapter 1 for a fuller description of the corpora used in this study.

First, however, this chapter addresses the methodological challenges posed by the actual data in the *NottDeuYTSch* corpus and how they had to be sampled and cleaned to ensure the accurate statistical analyses. The results of these analyses are presented in 5.2, and then further built upon in Chapter 6.

## 5.1 Sampling the *NottDeuYTSch* corpus for syntactic analysis

The general methods of constructing the corpus and cleaning the data have been covered in Chapter 3. This section outlines the challenges involved in carrying out quantitative analysis of paratactic clauses introduced by subordinating conjunctions. Firstly, it was not feasible to capture every clause in the *NottDeuYTSch* corpus with a paratactic structure that followed a subordinating conjunction. This would have required extensive data processing to account for every orthographic variation of every finite verb and subordinating conjunction. Instead, a stratified sample of the data was taken, described in Sections 5.1.1 and 5.1.2.

Pasch et al. (2003: 354) identified 69 different subordinating conjunctions, and the *DWDS-Kernkorpus 21* contains at least 8,500 different finite verb types. In addition, the large number of spelling variations makes it difficult to identify and correctly tag all relevant parts of speech automatically, in this case, all finite verbs and all subordinating conjunctions. Using a part-of-speech tagger would either have forced the analysis to rely on standardised spelling, which might mean that some paratactic syntagma are missed, or have required each token in the *NottDeuYTSch* corpus to be standardised, e.g. identifying and changing each non-standard spelling, which was not possible within the scope of the project. Additionally, relying on a part-of-speech tagger for all finite verbs would have led to too many false positives in the data, due to the large number of orthographic variants and homographs, and would have required extensive data cleaning. For example, *weiß*, is both the first and third person singular present form of *wissen* ('to know') and the colour white. The three part-of-

speech taggers tested in this study ('OpenNLP', 'qdap', and 'RDRPOSTagger') all had difficulty categorising ambiguous instances such as in Example 5.1, where human observers would be able to make an accurate decision:

> **Example 5.1** (2017)
> Ich würde mich freuen **wenn ich weiß oder rosa gewinne** ihr zwei seid die besten♡♡♡
> ('I would love it **if I win white or pink** you two are the best♡♡♡' rather')

Two common cases in the *NottDeuYTSch* corpus illustrate the problem of non-standard spellings and forms. Firstly, *ist*, the third person singular present indicative form of *sein*, is often written as *is*, using a graphical representation of the common spoken simplification of the consonant cluster *-st* (Durrell 2003: 17). Secondly, the final *-e* in some verb conjugations is often omitted, a written representation of schwa-apocope in spoken German, for example, writing *hab*, instead of *habe* for the first person singular present indicative form (Kohler 1995: 207; Kohler & Rodgers 2001: 14; Berend 2005: 157).

In the absence of adequate part of speech tagging to account for the large number of orthographic variations, Regular Expressions (RegEx) were used to identify and capture clauses containing subordinating conjunctions with a finite verb. Since separate RegEx patterns had to be written for each conjunction and verb, the analysis was necessarily limited to a manageable and representative sample which could be processed using RegEx.

As an example, the RegEx pattern used to search for the most common spelling variations and inflections of the verb *wollen* can be written as `\\bw[oi]l+[est]*n?\\b`. A full breakdown of the logic is explained step-by-step in Table 5.5.

Particular care needs to be taken with RegEx patterns, because they can only capture the patterns they are written to match, and so will not capture spelling variants that are not accounted for in the expression. Nor can they differentiate between identically written lexemes that are different parts of speech, so

Table 5.5: Explanation of an example Regular Expression

| Step | RegEx Symbol | Explanation |
|---|---|---|
| 1 | `\\b` | A word boundary, i.e. no alphabetical character precedes the first letter (in this case, <w>). This ensures that no past participles of 'wollen' are captured, e.g. 'gewollt' |
| 2 | `w` | Captures the letter <w> |
| 3 | `[oi]` | Captures one instance of either of the letters <o> or <i> |
| 4 | `l+` | Captures at least one instance of the letter <l>. This accounts for spelling variations e.g.: "ich **wil** es haben weil es einfach geil ist und ich das t-shirt will" (16.12.2012) "fett ich **willll** auchhhhh" (07.04.2017) |
| 5 | `[est]*` | Captures any number (including 0) of combinations of letters containing <e>, <s>, or <t>, for example the -<test> ending of 'wolltest' |
| 6 | `n?` | Captures 0 or 1 instance of the letter <n>. This is to capture 1st and 3rd person plural declensions of the verb |
| 7 | `\\b` | A word boundary, i.e. no alphabetical character succeeds the last letter (in this case, <e>, <n>, <s>, or <t>). This ensures that no present participles of 'wollen' are captured, e.g. 'willend' |

manual processing is required to remove these false positives. For example, spelling variants identical to another lexeme often pose problems, such as in the case of *dass* (the subordinating conjunction), for which the spelling *das* (the standard spelling of the phonologically identical neuter definite article) is often used. Similarly, for *seid* (the second person plural present tense form of the verb *sein*), the form *seit* (identical phonologically and in spelling to the preposition or conjunction) is often used. Restricting the range of potential variables was therefore paramount to search the *NottDeuYTSch* corpus efficiently for a maximally representative sample of examples of parataxis following subordinating conjunctions. The process for arriving at this representative sample is described in Sections 5.1.1 to 5.1.3 below.

### 5.1.1 Selecting subordinating conjunctions

The investigation used an adapted list of subordinating conjunctions from the *Handbuch der deutschen Konnektoren* (Pasch et al. 2003: 354) to investigate the different rates of parataxis. The original list of 69 different subordinating conjunctions and conjunction clusters was adapted from Pasch et al. (2003) in the following ways:

**1. Subordinating conjunctions identical with other frequently occurring parts of speech were excluded.** Fifteen subordinating conjunctions that are identical in form to other parts of speech were excluded: *als*, *bis*, *da*, *damit*, *ehe*, *gleichwohl*, *nun*, *seit* (although *seitdem* is included), *so*, *sosehr*, *soviel*, *statt*, *trotzdem*, *wie*, and *wo*. These posed too great a challenge for data collection, as they returned so many false positives, and sample studies of the *NottDeuYTSch* corpus revealed that these 15 lexical items were less common as subordinating conjunctions than other parts of speech. For example, there were, in fact, no occurrences of *statt* as a conjunction in the *NottDeuYTSch* corpus; all tokens were prepositions (meaning 'instead') or the particles of particle verbs, e.g. *stattfinden*. The methodological processes leading to the exclusion of these conjunctions is outlined below.

The potential conjunctions to exclude were cross-referenced with the *DeReKo* (*Deutsche Referenz Korpus*) and *DWDS WebXL Korpus*, comparing the frequencies as a subordinating conjunction and as another part-of-speech across similar time frames for consistency. If a lexical item occurred more frequently as another part of speech in all three corpora, rather than as a subordinating conjunction, it was excluded from the analysis. Figure 5.1 shows the results, with the shaded sections of the bars representing the instances of the lexical item used as a subordinating conjunction per million comments, whilst the solid sections represent the instances of the lexical item per million comments used as another part of speech. Nine of the lexical items (*als*, *bis*, *damit*, *nun*, *seit*, *so*, *statt*, *trotzdem*, and *wo*) were primarily or exclusively used as other parts of speech in both the *DWDS Webkorpus* and *DeReKo*. For example, *trotzdem* primarily occurred in three corpora as a conjunctive adverb ('Adverbkonnektor'), i.e. an

adverb that semantically links to the previous sentence or clause (Breindl 2008: 14-16), as shown in Example 5.2, although it is also found in the *NottDeuYTSch* corpus as a discourse marker (see Chapter 2, Section 2.2.4), rather than as a subordinating conjunction.

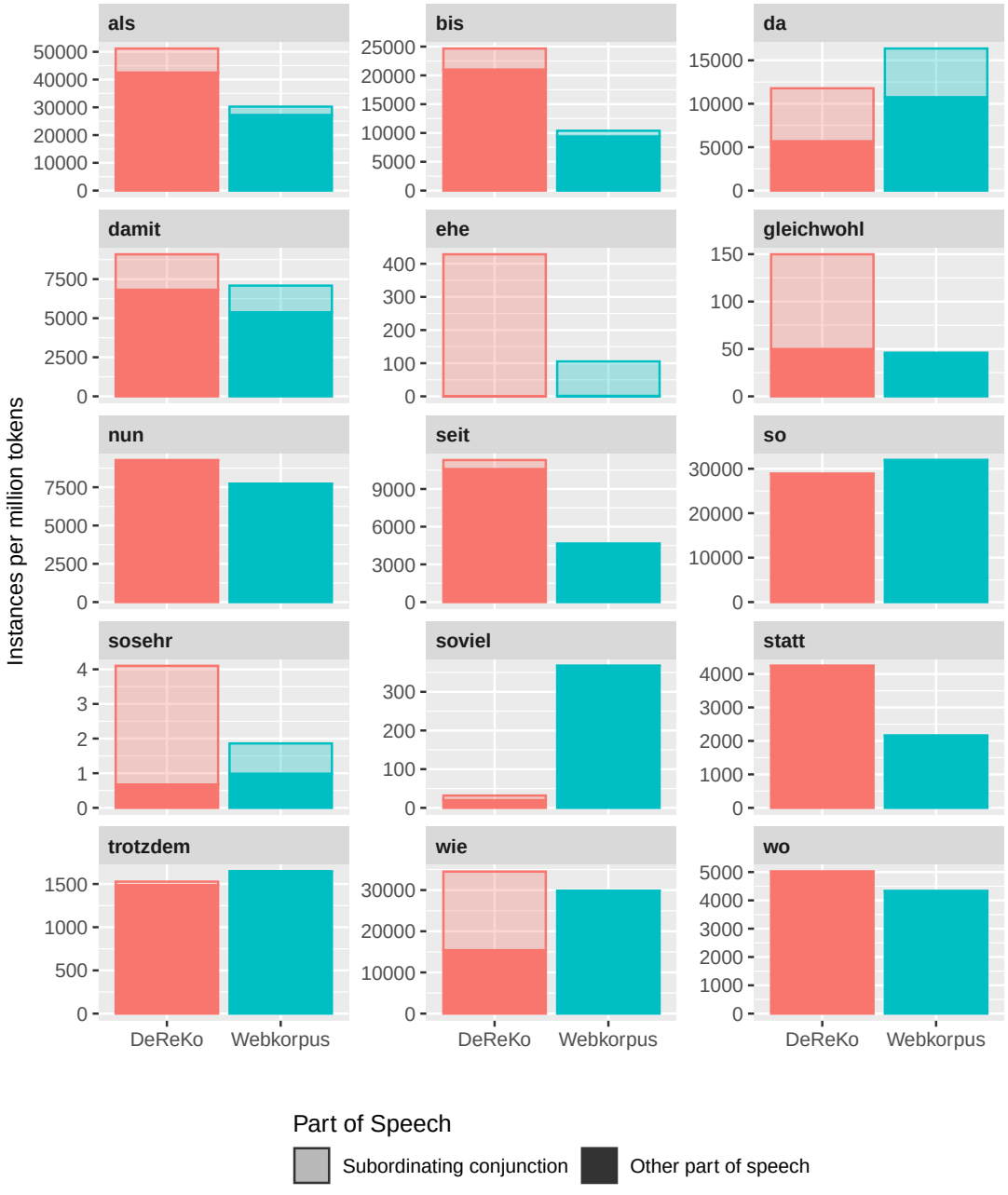



Figure 5.1: Frequency of part-of-speech classification of the excluded uncommon conjunctions in the *DeReKo* (red) and *DWDS Webkorpus* (blue)

**Example 5.2** (Conjunctive adverb, taken from Breindl, Volodina and Waßner (2014: 21))

> Nashörner wirken schwerfällig. **Trotzdem können sie sehr schnell laufen.**
> ('Rhinoceri seem sluggish. **Nevertheless they can run very fast.**')

Despite functioning as a discourse marker or conjunctive adverb in the data in the *NottDeuYTSch* corpus, *trotzdem* was captured by the RegEx patterns because it appeared in the position that would be occupied by a conjunction if it had paratactic syntax structure (as in Example 5.3 below, where *trotzdem* functions as a discourse marker).[2] *Trotzdem* was therefore excluded from consideration because it barely ever occurs in the corpus as a subordinating conjunction. However, occurrences of subordinating conjunctions, including *trotzdem*, possibly functioning as discourse markers are present throughout the *NottDeuYTSch* corpus, often followed by main clause syntax structure (see Freywald 2016), and they are discussed in Chapter 6, Section 6.2.

> **Example 5.3** (2015)
> [...] aber **trotzdem das ist nur ein Spiel** und die hater sollen sich nicht so aufregen
> ('[...] but **anyway this is only a game** and the haters shouldn't get so annoyed')

Five lexemes (*da*, *gleichwohl*, *sosehr*, *soviel*, and *wie*) showed different usage patterns in the *DWDS Webkorpus* and *DeReKo*. For example, in the *DeReKo*, *gleichwohl* is used primarily as a subordinating conjunction, but not in the *DWDS Webkorpus*, where it occurs 30 times more frequently as an adverb than as a conjunction. Only one of the 14 lexemes, *ehe*, was used primarily as a subordinating conjunction in both of the comparison corpora. However, in the *NottDeuYTSch* corpus, with its wide range of non-standard orthography, including the non-capitalisation of nouns, many more false positives are captured when searching for *ehe* than in the other corpora, as the patterns used to search the corpus have to allow for this variation. For example, *ehe* is often used as a noun (*die Ehe*, 'marriage') in the *NottDeuYTSch* corpus:

[2] If it were functioning as an adverb, the standard word order would be "trotzdem ist das [...]".

**Example 5.4** (2015)
#NetzFragtMerkel sind sie tatsächlich gegen die homo-**ehe** oder vertreten sie nur die kapitalistische meinung ihrer partei?
('#InternetAsksMerkel are you really against gay **marriage** or are you just representing the capitalist opinion of your party?')

In Example 5.4, the standard capitalisation of formal personal and possessive pronouns (*Sie*, *Ihrer*) and the nouns (*Homo-Ehe*, *Meinung*, and *Partei*) is lacking, and this includes *ehe* for *Ehe*. The form *ehe* also occurs in the *NottDeuYTSch* corpus as orthographic variation of the adverb *eher* and of *eh*, the modal particle (used originally in the Southern German/Austrian dialect region and now widespread throughout all regions, Durrell 2017: section 9.1.9). In Example 5.5, *Ehe* is used as a modal particle and is an orthographic variation of *eh*, both in spelling and capitalisation (possibly due to an autocorrect feature of mobile communication). Identifying and removing all of the false positives of *ehe* would have required too much manual data processing. Therefore, *ehe* was also excluded from the analysis.

**Example 5.5** (2016)
kriege ich **Ehe** nicht
('I don't get it **anyway**')

**2. Uncommon conjunctions were excluded.** Seven conjunctions listed in Pasch et al. (2003) were not analysed because they did not appear in the *NottDeuYTSch* corpus in clauses with the twenty verbs that were selected for the investigation (on the selection of verbs, see Section 5.1.2 below). A further three were excluded because they appeared fewer than five times. Their infrequency means that they cannot be investigated statistically.

**Conjunctions that did not appear with the verbs chosen for sampling:**

- *alldieweil*
- *derweilen*
- *obzwar*
- *sintemal(en)*

- *wenn zwar/wennzwar*
- *wiewohl*
- *wofern*

**Conjunctions that appeared fewer than five times in total:**

- *insoweit* (1)
- *obschon* (1)
- *wenngleich* (3)

**3. Conjunction clusters were grouped together.** All 42 *dass* clusters listed in Pasch et al. (2003), the *Wörterbuch der Konnektoren* (Leibniz-Institut für Deutsche Sprache 2018e), and Breindl, Volodina, & Waßner (2014: 17), e.g. *kaum dass*, were combined into a single category called "*dass* cluster", covering all clusters where *dass* is separated from the rest of the cluster by a comma, e.g. *vorausgesetzt, dass*. These were combined into one category so that they could still be statistically analysed. Similarly, *insofern* and *insofern als* were combined into one entry, as were *insoweit* and *insoweit als*. Five conjunction clusters containing *wenn* were combined into a single *wenn cluster* category. They are:

- *als wenn*
- *auch wenn*
- *selbst wenn*
- *sogar wenn*
- *wie wenn*

**4. The addition of *wobei*.** *Wobei* was added to the list of selected conjunctions. It is often included in research on non-standard parataxis, but Pasch et al. (2003) and Breindl, Volodina, & Waßner (2014) do not include it in their list of subordinating conjunctions. Instead they treat it as a *Postponierer* ('postpositioner'), a conjunction that requires verb-final word order but that is not technically a subordinating conjunction. Unlike a subordinating conjunction, which can come before the main clause ('*anteponiert*', e.g. *weil ich müde bin, muss ich schlafen*), a subordinating clause containing a *Postponierer* must always follow

the main clause (Leibniz-Institut für Deutsche Sprache 2018a). However, Durrell (2017: 17.3) does not make an explicit distinction, and includes *wobei* alongside subordinating conjunctions of time.

The decisions outlined above resulted in a list of 30 conjunctions (or rather: 28 conjunctions and two groups of conjunction clusters) included in the analysis: *als ob*, *anstatt*, *bevor*, *dass*, *dass* clusters, *derweilen*, *falls*, *im Falle*, *indem*, *indessen*, *insofern als*, *nachdem*, *obgleich*, *obwohl*, *seitdem*, *sobald*, *sofern*, *solange*, *sooft*, *soweit*, *sowie*, *während*, *währenddessen*, *weil*, *wenn*, *wenn auch*, *wenn cluster*, *wennschon*, *wobei*, and *zumal*. In total, ten of these 30 conjunctions introduced paratactic clauses as subordinating conjunctions in the *NottDeuYTSch* corpus but four of these were excluded from further analysis (*bevor*, a *dass* cluster, *obgleich*, and *wobei*). Three of these excluded conjunctions, *bevor* ('before'), a *dass* cluster, and *obgleich* ('although'), introduced parataxis, but the constructions were part of written verse, as shown in Example 5.6. These instances of stylised verse were excluded from analysis as the word order is deliberately non-standard to fit the rhyming scheme and do no reflect spontaneous language production.[3]

**Example 5.6** (2015)
Für [YOUTUBER] war's ein gutes Jahr // **obgleich** Verluste waren da
('For [YOUTUBER] it was a good year // **although** losses were here')

Furthermore, although the 'postpositioner' conjunction *wobei* has appeared in analyses of parataxis in previous scholarship alongside other subordinating conjunctions (e.g. Günthner 2002; Breindl, Volodina, & Waßner 2014; Freywald 2016), it is not attested as a conjunction in the *NottDeuYTSch* corpus when introducing paratactic clauses. Instead, *wobei* only occurs as a discourse marker in these cases (n = 23, 8.6 ipmc), as demonstrated in Example 5.7, where it can be interpreted as introducing additional information and a narrative sequence, two of the functions of a discourse marker, discussed in Chapter 2, Section 2.2.4 above. Therefore, paratactic *wobei*-clauses introduced by *wobei*

[3]Such rhymes were particularly common at Christmas, with channels explicitly encouraging attempts at poetry as part of their seasonal content, such as video advent calendars.

have been recorded in the 'ambiguous cases' column. The use of conjunctions as discourse markers is examined in greater detail in Chapter 6, Section 6.2, as part of an investigation into changes to the grammatical functions of subordinating conjunctions.

> **Example 5.7** (2017)
> Oh die Kamera in Rosa ist einfach der Hammer 💝🦄😍 **wobei** klassisch weiß geht immer 🙈
> ('Oh the pink camera is simply amazing **although** with classic white you can't go wrong :D')

Examples of the use of the six remaining conjunctions (*dass*, *obwohl*, *solange*, *weil*, *wenn*, and *wenn* clusters) are presented in Table 5.6.

Table 5.6: Example comments from the *NottDeuYTSch* corpus of paratactic word order following the selected conjunctions

| **Subordinating Conjunction** | **Example Comment from the *NottDeuYTSch* corpus** | **Year** |
|---|---|---|
| dass | Ich denke, **dass** der Baum hat 50,00€ gekostet | 2013 |
| obwohl | geh dich vergraben, **obwohl** du kannst dir ja keine schaufel leisten xD | 2014 |
| solange | Geschenke kannst du trotzdem bringen, **solange** es sind keine Kippen | 2015 |
| weil | macht doch mal bitte was lustiges **weil** das is iwie voll unlogisch und kindisch -.- | 2010 |
| wenn | Ich stelle mich vor guckt auf mein Kanal **wenn** ihr wollt Abo | 2015 |
| wenn cluster | Die Geschichten in diesen Spielen liebe ich **auch wenn** ich mache Sachen oder Entscheidungen komisch | 2017 |

## 5.1.2 Selecting 20 common verbs

In order to keep the scale of the investigation manageable, it was decided to limit analysis to clauses containing twenty common verbs. The twenty selected verbs accounted for 324,275 comments, which is 82.6% of all comments (n = 392,492) that contained one or more of the selected subordinating conjunctions. We can calculate the margin of error to determine the amount of random sampling error in the sample of the twenty selected verbs, compared to using the entire corpus (Moore, McCabe, & Craig 2017: 349-352). The formula for

calculating the margin of error (MoE), corrected for finite population, is shown in Equation (5.1). $Z_\gamma$ is the Z-score at a given confidence level, $\gamma$, (2.576), $\hat{p}$ is the sample proportion (0.826), $n$ is the sample size (324,275), and $N$ is the population size (392,492).[4]

$$\begin{aligned} MoE &= Z_\gamma \cdot \frac{\sqrt{\hat{p} \cdot (1-\hat{p})}}{\sqrt{\frac{(N-1)\cdot n}{(N-n)}}} \\ &= 2.576 \cdot \frac{\sqrt{0.826 \cdot (1-0.826)}}{\sqrt{\frac{(392492-1)\cdot 324275}{(392492-324275)}}} \\ &= 0.00071 \end{aligned} \tag{5.1}$$

At a 99% confidence level, the margin of error is 0.071%, or rather there is a 99.929% probability that the sample is large enough to be representative of the corpus as a whole. While fewer verbs could have been selected, the higher number of verbs provides a wider variety of examples for qualitative analysis, without requiring extensive coding to account for the many orthographic variations of verb conjugations.

To identify the twenty finite verbs for the analysis of the *NottDeuYTSch* corpus, fully tagged corpora from the *Digitales Wörterburch der deutschen Sprache* (DWDS) were used. As the *NottDeuYTSch* corpus is not fully tagged, I could not simply generate a token frequency list to select the verbs, as it is not possible to distinguish between the infinitive and certain finite forms, such as the first or third person plural (e.g. *haben*: *wir haben* and *zu haben*) without considerable extra data processing. Neither the token frequency list nor the DWDS corpora can distinguish between the uses of *haben*, *sein*, and *werden* as auxiliary and lexical verbs, which would have aided larger-scale analyses involving tense differences in paratactic clauses. Therefore, potential differences between verb usage as either auxiliary or lexical were not explored. Three different corpora hosted by the *DWDS* were selected, which covered potential longitudinal and

[4] For a fuller explanation of Z-scores and the margin of error, see Moore, McCabe, & Craig (2017: 349–52), and for finite population corrections, see Ramachandran & Tsokos (2020: 187).

register-based language differences:[5]

1. The *DWDS-Kernkorpus 21* (*DWDS-KK21*). As the corpus contains written language taken from newspapers, scientific and functional texts, and poetry and fiction from the period 2000-2010, it overlaps with the time period of the *NottDeuYTSch* (2008-2018), and functions as the control, representing standard written German.
2. The *DWDS Webkorpus 2016c* (*DWDS-WK*). This corpus spans the time period closest to the *NottDeuYTSch* corpus and also consists entirely of online sources. It contains online German language taken from 8.2m webpages from professional, hobby and organisational German-language websites in the time period 2001-2016. To match the time period of the *NottDeuYTSch* corpus as closely as possible, I restricted the search to the period 2008 (the first year of the data in the *NottDeuYTSch* corpus) to 2016 (the latest year of data in the *Webkorpus*).
3. The *Dortmunder Chat-Korpus* (*DCK*). As this corpus was constructed from a range of online chats, e.g. casual conversations, semi-structured chats from university students and online question-and-answer sessions from politicians and celebrities, it is closest in style to the *NottDeuYTSch* corpus, although it uses the oldest data sources (1998-2006).

The following search term was used to identify finite verbs in the *DWDS* corpora.

```
COUNT( $p=/V\wFIN/#ASC_DATE[2008,2016] )
#BY[$l] #DESC_COUNT
```

This search term searches the corpus using part-of-speech tags from the Stuttgart-Tübingen Tagset (STTS). I searched for the number of hits (`COUNT()`) of all finite verbs (`$p=/V\wFIN/`). The forward slashes allow the insertion of a Regular Expression, `V\wFIN/`, which matches the following multiple relevant part-of-speech tags:

1. Finite modal verbs (`$p=VMFIN`)
2. Finite HSW (*haben*, *sein*, and *werden*) verbs (`$p=VAFIN`)[6]

[5] A fuller description of the corpora is found in Chapter 1.

[6] These verbs are also finite auxiliary verbs in perfect, pluperfect, and future tenses, but the

3. Finite full verbs (`$p=VVFIN`)

The search was restricted to the time period 2008-2016 (`#ASC_DATE[2008, 2016]`), and the results were grouped by lemma (`#BY[$l]`) and sorted by highest to lowest frequency (`#DESC_COUNT`).

This returned a table of 1,000 verb types (the maximum number that can be returned by a search query) ranked in order of frequency, from which I took the top 20 results. Table 5.7 lists the top 20 verbs appearing in finite forms in each of the three corpus searches. There is considerable overlap between the results for the three corpora; the verbs that only appear in the top twenty for one corpus are listed in bold.

Table 5.7: The 20 most frequent verbs (finite forms) from the selected *DWDS* corpora and the 20 selected verbs for analysis of the *NottDeuYTSch* corpus

| # | DWDS-WebXL Korpus | DWDS-Kernkorpus 21 | Dortmunder Chat-Korpus | Selected Verbs |
|---|---|---|---|---|
| 1 | finden | **bieten** | **denken** | denken |
| 2 | geben | **dürfen** | finden | finden |
| 3 | gehen | finden | geben | geben |
| 4 | haben | geben | gehen | gehen |
| 5 | kommen | gehen | **glauben** | glauben |
| 6 | können | haben | haben | haben |
| 7 | lassen | kommen | kommen | kommen |
| 8 | liegen | können | können | können |
| 9 | machen | lassen | machen | lassen |
| 10 | mögen | liegen | **meinen** | machen |
| 11 | müssen | machen | mögen | mögen |
| 12 | sagen | müssen | müssen | müssen |
| 13 | sehen | sagen | sagen | sagen |
| 14 | sein | sehen | sehen | sehen |
| 15 | sollen | sein | sein | sein |
| 16 | stehen | sollen | sollen | sollen |
| 17 | werden | stehen | stehen | stehen |
| 18 | wissen | werden | werden | werden |
| 19 | wollen | wissen | wissen | wissen |
| 20 | **wünschen** | wollen | wollen | wollen |

The three corpora share seventeen of the twenty most frequent verbs occurring in finite forms, which were therefore selected to analyse the *NottDeuYTSch* cor-

DWDS corpora do not differentiate when searching between their use as full or auxiliary verbs. For this reason, the chapter uses the deliberately underspecified label HSW (named after the three captured verb forms, *haben*, *sein*, and *werden*).

pus. The other most common verbs in the corpora differed, with the differences reflecting their varying communicative contexts. In the *DWDS-Kernkorpus 21*, the higher frequency of *bieten* (551 instances per million comments - ipmc) is typical of the more formal style of the corpus data that often leads to more complex verb constructions, shown in Example 5.8.

> **Example 5.8** (Jörg Lau (2015), DWDS-KK21 2000/zeit2000_13_48_1812)
> Die wahrscheinliche Übernahme durch den Bund und der Wechsel an der Spitze **bieten** jetzt die Gelegenheit, Inventur zu machen und die Aufgaben einer Einrichtung neu zu bedenken, in der manche schon ein Überbleibsel des Kalten Krieges sehen, auf das man heute gut und gerne verzichten könnte.
> ('The likely federal takeover and the change at the top now **offer** an opportunity to take stock and rethink the tasks of an institution that some already see as a remnant of the Cold War that could well be dispensed with today')

As the *NottDeuYTSch* corpus has a less formal style than the *Kernkorpus 21*, *bieten* was not selected. The *NottDeuYTSch* corpus has more in common with the *Chat-Korpus* and the *DWDS Webkorpus*: the *Chat-Korpus*, comprised of informal online conversations, contains language that is closer in register and structure (for example, the two corpora have similar average sentence length of just over eight tokens per sentence), and the *Webkorpus* spans a very similar time frame of online language to the *NottDeuYTSch* corpus, although it contains more formal documents and structures. In both of these corpora, especially in the *Chat-Korpus*, there were higher frequencies of verbs that express opinions (sometimes referred to as mental verbs or mental content verbs – MCVs), e.g. in the field of language acquisition: Shatz, Wellman, & Silber 1983; Papafragou, Cassidy, & Gleitman 2007), including *denken* ('to think'; 10,187 instances per million comments), *glauben* ('to believe'; 10,195 ipmc), *meinen* ('to mean/opine'; 7,042 ipmc), and *wissen* ('to know'; 8,473 ipmc). These verbs are characteristic of a more informal style, where personal reflection is frequent, unlike the more formal, impersonal styles of published writing in the *DWDS Kernkorpus*

*21*. These verbs are, accordingly, more commonly preceded by personal pronouns than noun phrases, as shown in Examples 5.9 to 5.11, in particular the first person singular pronoun, *ich*, which introduces personal reflection. In all the corpora, *ich* was the most frequent personal pronoun with every MCV (such as in Example 5.9), but the usage of *ich* with *meinen* is considerably lower than with the other MCVs. *Meinen* is more frequently used with the informal second person singular, *du*,[7] as illustrated in Example 5.10 from the *Chat-Korpus* below.

**Example 5.9** (Unknown (2003), DCK doc_p2_281)
Weil ich **glaub** lance ist nicht in topform !!!
('Because I **think** lance is not in top form')

**Example 5.10** (Unknown (2003), DCK doc_p1_163)
find ich zwar nicht ..... aber wenn du **meinst**
('I don't **think** so ..... but if you say so')

**Example 5.11** (Unknown (2003), DCK doc_p1_168)
[_PERSONNAME-26_] ist norwegisch ... und ist jetzt mein 2.vorname , den jeder [_PERSONNAME-28_] schreibt , weil er **denkt** er hätte sich verhört oder verlesen * grmpf *
('PERSONNAME 26 is Norwegian ... and is now my middle name , which everyone writes as PERSONNAME 28, because they **think** they misheard or misread * hmph *')

We also see examples of clauses solely containing a subordinating conjunction followed by a personal pronoun and then a mental content verb ("weil ich glaub", "wenn du meinst"). This structure is very typical for these verbs, but is ambiguous as to whether the finite verb is in the second or final position, and is therefore categorised as an 'ambiguous case', as explained in Section 5.1.3.2. Most often these clauses fulfil a hedging function. If we take Example 5.9, the commenter does not wish to state outright that Lance is not in top form, so hedges it as their personal opinion. Figure 5.2 shows that MCVs are

[7] The use of *meinen* in the second person singular (*meinst*) was around 10 times higher compared to other MCVs in the *Chat-Korpus*, 4 times higher in the *Kernkorpus 21*, and 1.5 times higher in the *NottDeuYTSch* corpus.

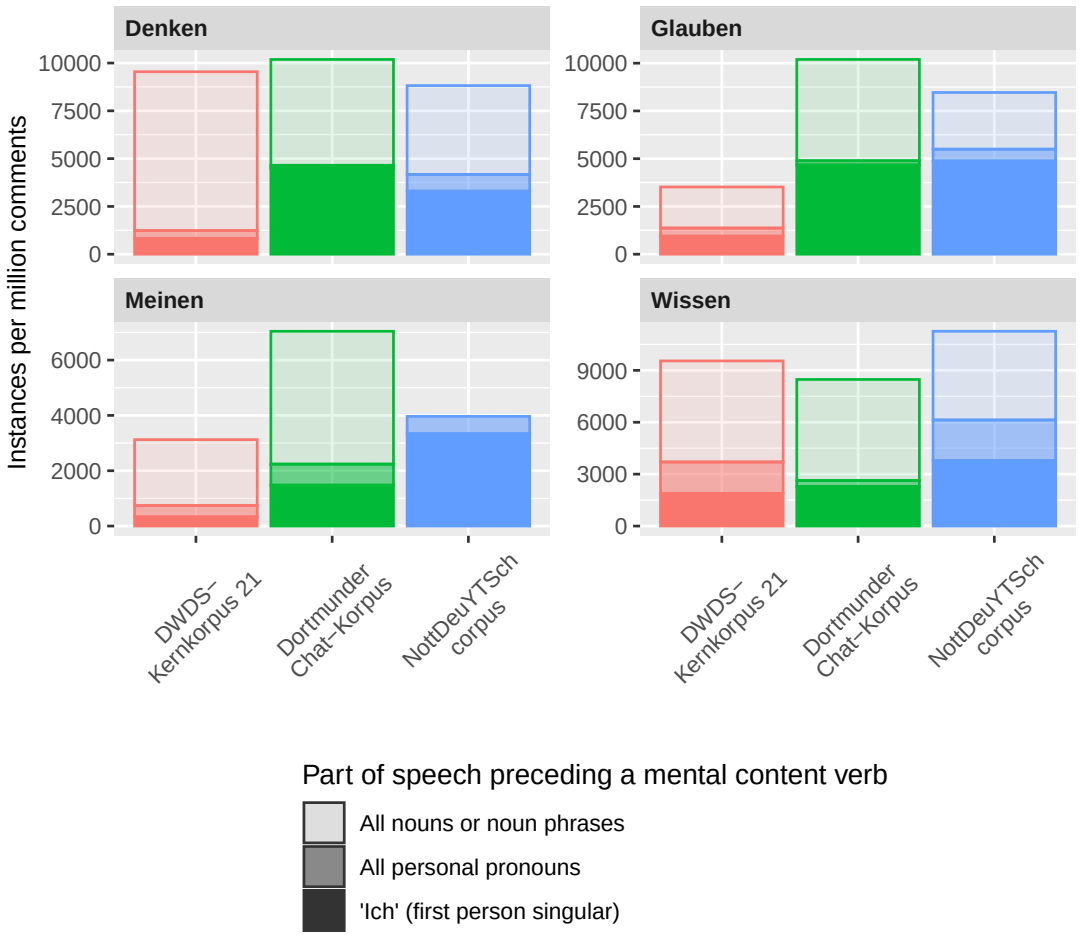
far more frequent overall in the *Dortmunder Chat-Korpus* and *NottDeuYTSch* corpus than in the *DWDS Kernkorpus 21*.[8] Given the high frequency of these MCVs in the corpora most similar in style to the *NottDeuYTSch* corpus, *denken* and *glauben* were selected to analyse clause structure following subordinating conjunctions along with the seventeen frequent verbs common to each corpus, and *lassen*, the most frequent other verb when the frequencies of the *Dortmunder Chat-Korpus*, *DWDS Kernkorpus*, and *DWDS Webkorpus* were combined.[9] The 20 selected verbs, as listed in the final column of Table 5.7, are *denken*, *finden*, *geben*, *gehen*, *glauben*, *haben*, *kommen*, *können*, *lassen*, *machen*, *mögen*, *müssen*, *sagen*, *sehen*, *sein*, *sollen*, *stehen*, *werden*, *wissen*, and *wollen*.

### 5.1.3 Capturing paratactic clauses in the *NottDeuYTSch* corpus

Having restricted the corpus to clauses including the twenty selected verbs, sub-corpora of all the comments with each of the six qualifying subordinating conjunctions or clusters were created and were then processed to identify potentially paratactic clauses. This was initially done in five passes. The first pass captured clauses where the verb was the next token following the subordinating conjunction (so-called V1 (verb-first) parataxis, cf. Reis 2013: 224). The second and third passes captured clauses with one token between the conjunction and verb, i.e. potentially paratactic clauses with first fields of one token length, and then two tokens between conjunction and verb. The fourth and fifth passes, which searched for clauses with three and four tokens between conjunction and verb, returned no potentially paratactic clauses and so it was determined that no more passes would be made. To extend the distance between tokens would have risked capturing verbs in subsequent clauses, as often clauses are

[8]The frequency of *wissen* in the *NottDeuYTSch* corpus statistics has been adjusted to exclude *weiß* used as a colour, which was done by finding the difference between the total captures (n = 31,793) and the instances of *weiß* in the same clause as either the first or third person singular pronouns (n = 23,180).

[9]While *meinen* was not one of the three most frequent verbs in the combined corpus, it is worth nothing that *meinen* would, in any case, have presented difficulties for analysis, given the orthographic similarities between conjugations of the verb and the first person singluar possessive pronoun, *mein*, such as *ich meine* and *meine Hausaufgabe*, which would be difficult to handle using RegEx patterns in an untagged corpus.




Note: No data can be provided for 'meinen' preceding a noun or noun phrase in the 'NottDeuYTSch' corpus as the RegEx pattern search cannot distinguish between use as a verb and the orthographical similarities to the first person possessive pronoun, 'mein'.

Figure 5.2: A comparison of personal pronoun frequency with mental content verbs in the *DWDS Kernkorpus 21*, *Dortmunder Chat-Korpus* and *NottDeuYTSch* corpus

not separated by punctuation. Each sub-corpus was manually checked and iteratively processed using a tailored list of stop words for further removal of false positives, before a final complete manual check of all captured clauses. Clauses were sorted into three categories, each of which is discussed in more detail below:

1. 'False Positives' (cases that were captured as seemingly paratactic and which had to be manually removed, see Section 5.1.3.1 below)
2. 'Ambiguous' cases (cases where it was not possible to discern if the clause was paratactic or hypotactic, see Section 5.1.3.2 below)
3. 'Parataxis' (paratactic word order following subordinating conjunctions, as analysed statistically and longitudinally in Section 5.2, and with regard to potential grammatical processes in Chapter 6)

#### 5.1.3.1 Technical false positives

The initial RegEx patterns used to search for paratactic word order captured many false positives, which fall into seven main groups:

**1. Alternative part of speech:** This group encompasses the use of a potential subordinating conjunction as a different part of speech, e.g. in Example 5.12, where *währendessen* functions as an adverb followed by a non-standard periphrastic verb construction, *ist ausrechnen*, with the lexical verb, *rechnen* in the first field followed by the finite auxiliary verb, *ist* in the left bracket, i.e. the position of the finite verb in a main clause.

> **Example 5.12** (2018)
> Also zuerst muss Lucy wieder kacken gehen und **währenddessen rechnen ist Ludgar eine mathematische Formel aus** wie man Leute rettet […]
> ('So first Lucy has to go for a poo again and **meanwhile Ludgar is calculating a mathematical formula** how to save people […]')

One potential way of avoiding similar false positives would be to account for the past participle between the (potential) subordinating conjunction and the

finite verb in the RegEx patterns. However, it was not possible to produce a pattern, e.g. `ge\\p{L}+(t|en)`, that would only capture past participles and not other parts of speech that match the pattern, e.g. the word *Geist*. Furthermore, this method would have returned different false positives, such as clauses containing ellipsis, where the article or pronoun is dropped, e.g. *weil baum ist* ('because [it] is [a] tree', 2017).

**2. Ellipsis:** This group consists of the omission of words from a clause, particularly pronoun dropping or the agentless passive, which means a finite verb is wrongly captured as being in second position, as shown in bold in Example 5.13.

> **Example 5.13** (2014)
> Das ist doch klar, dass es irgendwann zum Ende kommt **wenn getrollt wird**, weil es irgendwann immer einen gibt, der zu weit geht :D
> ('It is clear that at some point it will come to an end **if trolled [if there is trolling going on]**, because at some point there is always someone who goes too far :D')

Ellipsis and other economical uses of language are well documented as features of informal language (Durrell 2003: 37), online language (Siever 2006: 83) and youth language (particularly in settings influenced by Turkish, Bahlo & Klein 2017: 142-144), all of which are applicable to the *NottDeuYTSch* corpus.

**3. Cliticisation:** This group consists of constructions where a clitic is attached to a lexeme. Fagan (2009: 283) defines a clitic as "an unstressed word that cannot stand alone but must be attached to a neighboring stressed word, with which it forms a unit". For example, when the third person singular personal pronoun, *es*, is cliticised, it appears as *s* attached to the end of a word, e.g. *weil* becomes *weils* (as in Example 5.14), resulting in the verb appearing to the RegEx as if it were in the second position.

> **Example 5.14** (2017)
> [USERNAME] **weils** scheiße is

('[USERNAME] **because it**s shit')

This most often occurred with *wenn* and *weil* (see Example 5.14), or with the clitic bound to the first field constituent, such as a personal pronoun, e.g. *ichs*, for *ich es* (see Example 5.15). This is another example of schwa-apocope (as mentioned in Section 5.1), which results in the combination of the two words, often appearing in more formal environments as *wenn's* (which is accounted for in the RegEx). There were also a few examples of *wennst*, as shown in Example 5.15, where the [st] is a cliticisation of the second person singluar personal pronoun, *du*, i.e. a variation of *wenn du*, which is a feature of the Upper German language area (Bayer 1984).

**Example 5.15** (2017)
#fail hau mal wieder eine Chatroulette folge raus, natürlich nur **wennst** zeit hast
('#fail chuck a Chatroulette episode out again, obviously only **if you** have time')

**4. Non-standard orthographic practices:** The groups encompasses alternative spellings, including the use of spaces. These can result in a token being wrongly captured as a subordinating conjunction, such as writing *so* and another adverb or adjective as one word, instead of separately, e.g. *soweit* instead of *so weit*, as in Example 5.16. The six examples of this found in the *NottDeuYTSch* corpus were *sobald*, *sofern*, *solange*, *sooft* (Example 5.17), *soweit*, and *sowie*).

**Example 5.16** (2010)
bei 0:43 sieht man ihren Schlüpper, **soweit** habt ihr mich gebracht!!! :'(
('at 0:43 you can see her underwear, you have brought me **so far**!!! :'( ')

**Example 5.17** (2012)
[...] Und da ich meine Beste freundin (aus [PLACENAME] die stadt kennst du sicher) schon ein halbes jahr nicht mehr gesehen hab weil ich weggezogen bin und wir uns deshalb nicht **sooft** treffen können würde ich mit

ihr gerne ins [PLACENAME2] :)
('And since I haven't seen my best friend (from [PLACENAME] you know the city for sure) for half a year because I moved away and we can't meet **so often** I would like to take her to [PLACENAME2] :)')

**5. Complex verb constructions:** This group encompasses the use of three or more consecutive verb forms, such as perfect tense modal or double modal verb constructions in a subordinate clause, where the finite auxiliary verb is found at the front of the verbal construction, as demonstrated with "hatte kommen können" in Example 5.18 (for further reference, see Examples 15 and 16 in Leibniz-Institut für Deutsche Sprache 2018d).[10]

**Example 5.18** (2014)
Sie kamen nach draußen und fanden mich bevor ein Zug **hatte kommen können**.
('They came outside and found me before a train **had a chance to come**')

**6. Exbraciation (*Ausklammerung*):** Exbraciation is a standard grammatical feature in German, where a phrase is placed in the end-field (*Nachfeld*) of the sentence (highlighted in bold in Example 5.19), instead of in the midfield, which leads to it being captured by the RegEx.

**Example 5.19** (2011)
Das bei 0:57 war gar kein Creeper,weil der Arme hatte **und zwei Beine**
('That [thing] at 0:57 wasn't a creeper,because he had arms **and two legs**')

Exbraciation can occur regardless of phrase length. In Example 5.20, the lengthy phrase *von Leichenteilen und Foltertechniken und von der desolaten moralischen Verfassung seiner Brüder* is exbraciated, whilst in Example 5.21, it is only *in Berlin* that follows the verb *ankam*, instead of occurring at the end of the clause (*als sie in Berlin ankam*). Exbraciation tended to occur more frequently in corpora of formal language, such as the *DWDS Kernkorpus 21*, rather than corpora con-

[10] In fact, Variantengrammatik des Standarddeutschen (2020) demonstrates than the finite verb occurs in any position within a complex verb construction, i.e. at the front (***hatte*** *kommen können*), in the middle (*kommen* ***hatte*** *können*), or and the end (*kommen können* ***hatte***).

taining significant colloquial language, such as the *Dortmunder Chat-Korpus* and the *NottDeuYTSch* corpus.

> **Example 5.20** ( J. Buchsteiner (2000), DWDS-KK21 2000/zeit2000_18_8_2892 )
> Wenn er berichtet **von Leichenteilen und Foltertechniken und von der desolaten moralischen Verfassung seiner Brüder**, jetzt, wo der Krieg verloren scheint.
> ('When he tells **of body parts and torture techniques and of the desolate moral condition of his brothers**, now that the war seems lost.')

> **Example 5.21** (H. Müller (2003) DWDS-KK21 2003/Mueller_Koenig_2003)
> Sie hatte mich bereits verraten, als sie ankam **in Berlin**, und während sie ihren Verrat gestand, behauptet, sie könne nie etwas tun, was mir schadet.
> ('She had already betrayed me when she arrived **in Berlin**, and while confessing her betrayal, she claimed she could never do anything to harm me.')

**7. Homographs:** This groups consists of instances where an infinitive was orthographically identical to other conjugations of a verb, such as the first or third person plural present tense forms, e.g. *kommen* in Example 5.22. These had to be manually removed.

> **Example 5.22** (2017)
> [YOUTUBER]😂😁🚘🍘🍘 und hier **wenn du kommen kannst** dann dann da [YOUTUBER] also kannst du vielleicht morgen kommen [...]
> ('[YOUTUBER]😂😁🚘🍘🍘 and here **if you can come** then then here [YOUTUBER] so can you perhaps come tomorrow [...]')

#### 5.1.3.2 Ambiguous syntactical structure

As noted above, a notable minority of structures identified by the RegEx patterns were syntactically ambiguous. Ambiguous syntactic structures followed subordinating conjunctions in 4,533 comments (1.40%) of the 324,275 com-

ments that contained the 20 selected verbs. In these cases, it was not possible to establish whether the clause was hypotactic (the finite verb was in the final position) or paratactic (the finite verb directly followed the subject).

Ambiguous clauses typically have the structure `[SUBORDINATING CONJUNCTION] + [NOUN PHRASE] + [FINITE VERB]`, and it was very common that the ambiguous clauses were comprised of a subordinating conjunction (or conjunction cluster) followed by a personal pronoun and a mental content verb, as in Examples 5.23 and 5.24. As explained in Section 5.1.2 above, the high frequency of these structures reflects the personal content of the comments, i.e. more closely aligned with informal speech than formal written language. Of the 30 conjunctions initially examined, clauses with conditional or temporal conjunctions produced the highest rates of ambiguous clauses. There are three conditional conjunctions (*falls*, *in Falle*, *sofern*), and ten temporal conjunctions (*bevor*, *indem*, *nachdem*, *seitdem*, *sobald*, *solange*, *sooft*, *sowie*, *während*, and *währendessen*), as well as *wenn*, which can be used in both a conditional (Example 5.23) and a temporal sense (Example 5.24).

> **Example 5.23** (Conditional: 2017)
> ey an alle hater verpisst euch einfach **wenn ihr meint** er kann nicht singen dann ignoriert es einfach und machst besser **wenn ihr denkt** er kann nicht singen 😑
> ('hey to all the haters just fuck off **if you think** he can't sing then just ignore it and do better **if you think** he can't sing 😑')

> **Example 5.24** (Temporal: 2016)
> Und was darf's noch sein ein Interview mit pewdipie , nen Date mit [YOUTUBER] oder sowas Tut dein Kopf schlimm weh **wenn du denkst** ????
> ('And what else would you like an interview with pewdipie [a world-famous YouTuber], a date with [YOUTUBER] or something Does your head hurt **when you think** ????')

*Wenn* clauses are often inserted into the midfield of other clauses as am 'in-

dependent communicative unit' ('selbständige kommunikative Minimaleinheit,' Zifonun 1987),[11] i.e. a self-contained unit that is syntactically complete, if not semantically. The insertion of these phrases and syntactic detachment from the surrounding sentence is referred to by Pasch et al. (2003: 392–93) as "prosodically manifest disintegration", as shown in Example 5.25. In this example, the insertion of *wenn du willst* has a hedging, concessive effect.

**Example 5.25** (2015)
achja ich würde **wenn du willst** eine monster abwehr maschine bauen
('ah yes I would **if you want to call it that** build a monster defence machine')

Auer (1998: 284) describes these short constructions as "dependent main clauses" (*abhängige Hauptsätze*) following Müller (1971), and notes that they fall on continuum "between parataxis and hypotaxis" (Auer 1998: 297-298). Auer remarks that some such structures, such as *wenn du willst* above, can undergo grammaticalization so that the fixed phrase function as a discourse marker. This has been corroborated by later work on discourse markers containing the verbs *meinen* (Günthner & Imo 2003) and *glauben* (Imo 2011). Because it is not possible to definitively categorise these structures as either paratactic or hypotactic in the *NottDeuYTSch* corpus, they were classified as "ambiguous".

## 5.2 Analysing subordinating conjunctions and parataxis

After eliminating the false positives and ambiguous structures from the data sample, we were left with 5,404 paratactic structures that follow six of the 30 selected conjunctions (i.e. *dass*, *obwohl*, *solange*, *weil*, *wenn*, and *wenn* clusters).[12] These paratactic clauses following subordinating conjunctions are analysed in

[11] These are also referred to as turn-construction units in the field of discourse and conversation analysis (see Sacks, Schegloff, & Jefferson 1974: 721-722).
[12] As mentioned in Section 5.1.1 above, 20 conjunctions did not introduce any paratactic clauses, and four conjunctions were excluded from further analysis.

three stages. First, Section 5.2.1 focuses on the data in the *NottDeuYTSch* corpus, analysing the frequencies of paratactic clauses introduced by the six conjunctions, demonstrating that, compared to previous scholarship, the range of conjunctions that can be followed by parataxis has increased. Second, Section 5.2.2 presents a longitudinal investigation of the change in the frequency of this phenomenon over the course of the corpus, providing concrete empirical data conclusively demonstrating an increase in the overall usage of parataxis after subordinating conjunctions from 2008-2018. Third, Section 5.2.3 demonstrates that the phenomenon of paratactic syntax structures following subordinating conjunctions is far more prevalent in digital youth language than other registers of written language by comparing the data in the *NottDeuYTSch* corpus with other corpora. The comparative study quantitatively establishes that the frequency of paratactic clauses following subordinating conjunctions forms a continuum based on the acceptability of non-standard language use based on the communicative environment.

### 5.2.1 Parataxis in the *NottDeuYTSch* corpus

In response to claims made in previous scholarship that non-standard paratactic structures only follow causal or concessive subordinating conjunctions (e.g. Antomo & Steinbach 2010; Reis 2013), this Section briefly considers the frequencies of paratactic clauses introduced by the selected conjunctions in the *NottDeuYTSch* corpus, and the changes in frequency of this phenomenon over time. Table 5.8 shows the rates of parataxis and numbers of ambiguous cases for all 30 selected subordinating conjunctions, following the data cleaning processes to identify false positives and ambiguous cases in Sections 5.1.3.1 and 5.1.3.2.

As explained in Section 5.1.1 above, six conjunctions qualify for further analysis in this section: *dass* ('that'), *obwohl* ('although'), *solange* ('as long as/whilst'), *weil* ('because'), *wenn* ('when/if'), and *wenn* clusters. Examples of parataxis in clauses following these conjunctions are listed in Table 5.9 (repeat of Table 5.6).

Table 5.8: Frequency of the selected conjunctions that introduce paratactic clauses in the *NottDeuYTSch* corpus

| **Subordinating conjunction** | **Total number of clauses (with common verb)** | **Ambiguous cases** | **Paratactic clauses** | **Percent of clauses with parataxis** |
|---|---|---|---|---|
| als ob | 2 999 | 24 | 0 | 0.00 |
| anstatt | 1 714 | 0 | 0 | 0.00 |
| bevor | 2 994 | 37 | 0* | 0.00 |
| dass | 61 955 | 574 | 41 | 0.07 |
| dass cluster | 2 117 | 29 | 0* | 0.00 |
| derweilen | 27 | 0 | 0 | 0.00 |
| falls | 4 484 | 55 | 0 | 0.00 |
| im Falle | 84 | 0 | 0 | 0.00 |
| indem | 1 128 | 2 | 0 | 0.00 |
| indessen | 5 | 2 | 0 | 0.00 |
| insofern als | 80 | 0 | 0 | 0.00 |
| nachdem | 1 425 | 3 | 0 | 0.00 |
| obgleich | 10 | 0 | 0* | 0.00 |
| obwohl | 6 201 | 25 | 51 | 0.82 |
| seitdem | 907 | 0 | 0 | 0.00 |
| sobald | 714 | 5 | 0 | 0.00 |
| sofern | 187 | 5 | 0 | 0.00 |
| solange | 1 757 | 13 | 1 | 0.06 |

Table 5.8: Frequency of the selected conjunctions that introduce paratactic clauses in the *NottDeuYTSch* corpus *(continued)*

| **Subordinating conjunction** | **Total number of clauses (with common verb)** | **Ambiguous cases** | **Paratactic clauses** | **Percent of clauses with parataxis** |
|---|---|---|---|---|
| sooft | 22 | 2 | 0 | 0.00 |
| soweit | 1 120 | 14 | 0 | 0.00 |
| sowie | 1 091 | 1 | 0 | 0.00 |
| während | 2 136 | 6 | 0 | 0.00 |
| währenddessen | 109 | 0 | 0 | 0.00 |
| weil | 102 692 | 795 | 5249 | 5.11 |
| wenn | 119 303 | 2 894 | 58 | 0.05 |
| wenn auch | 256 | 0 | 0 | 0.00 |
| wenn cluster | 8 486 | 47 | 4* | 0.05 |
| wennschon | 5 | 0 | 0 | 0.00 |
| wobei | 800 | 23 | 0 | 0.00 |
| zumal | 267 | 0 | 0 | 0.00 |
| **Total** | **325 075** | **4 556** | **5404** | **1.66** |

* Paratactic clauses from text stylised as verse excluded from the table statistics

Table 5.9: Example comments from the *NottDeuYTSch* corpus of paratactic word order following the selected conjunctions

| Subordinating Conjunction | Example Comment from the *NottDeuYTSch* corpus | Year |
|---|---|---|
| dass | Ich denke, **dass** der Baum hat 50,00€ gekostet | 2013 |
| obwohl | geh dich vergraben, **obwohl** du kannst dir ja keine schaufel leisten xD | 2014 |
| solange | Geschenke kannst du trotzdem bringen, **solange** es sind keine Kippen | 2015 |
| weil | macht doch mal bitte was lustiges **weil** das is iwie voll unlogisch und kindisch -.- | 2010 |
| wenn | Ich stelle mich vor guckt auf mein Kanal **wenn** ihr wollt Abo | 2015 |
| wenn cluster | Die Geschichten in diesen Spielen liebe ich **auch wenn** ich mache Sachen oder Entscheidungen komisch | 2017 |

Of the 5,404 identified paratactic clauses, *weil* comprises 97.1% of instances (n = 5,249, 1,667 instances per million comments). *Weil* is also the conjunction with the highest proportion of paratactic clauses of all clauses introduced by a conjunction (5.1% of all *weil* clauses in the corpus were paratactic). This provides a quantitative justification for the prominence of *weil* clauses in previous research. However, paratactic *obwohl* clauses occur far less frequently (n = 51, 16.2 ipmc), despite often featuring alongside *weil* in scholarship. These frequencies directly contradict the observation of Schäfer & Sayatz (2016: 245) that "obwohl-V2 and weil-V2 occur with almost equal frequency" in the *DECOW12Q* corpus, a corpus of 1.8 billion tokens containing "almost exclusively forum discussions and blogs" (Schäfer & Sayatz 2016: 227), i.e. informal digital writing similar to the *NottDeuYTSch* corpus, although the age ranges of the contributors is not known. Potential reasons for this difference in frequency are suggested in Section 5.2.3 below.

Strikingly, paratactic *dass* clauses (n = 41, 13.0 ipmc) occur in the *NottDeuYTSch* corpus almost as frequently as with *obwohl*, yet these have not seen the same level of analysis in research (see Freywald & Simon 2007; Freywald 2016), and similarly, paratactic *wenn* clauses (n = 58, 18.4 ipmc) have not been covered in previous studies, despite occurring in the *NottDeuYTSch* corpus at similar frequencies. The lack of coverage in previous scholarship can be partially explained by the far greater proportion of *obwohl* clauses (0.82%) compared to

*dass* (0.07%) and *wenn* (0.05%) clauses that are paratactic in the *NottDeuYTSch* corpus (a significant difference in both cases, $\chi^2$, $p < 0.01$).

Perhaps concessiveness, similar to that of *obwohl*, may also explain the use of parataxis in *wenn* and *wenn* cluster clauses in the *NottDeuYTSch* corpus. In *wenn* clauses that are semantically conditional, i.e. implying a relationship between a condition and a consequence (rather than temporal, establishing the chronological order of events), there is often an unmarked concessiveness, similar to *wenngleich*, *wennzwar*, or *wennschon*. This is demonstrated in Examples 5.26 and 5.27, where the commenters acknowledge that a condition will not be fulfilled: "though I do not have 100 nike air" (Example 5.26) and "although it is not good for the PC" (Examples 5.27).

**Example 5.26** (2016)
wäre besser **wenn ich habe** 100 nike air [a popular shoe]
('would be better **if I have** 100 nike air [a popular shoe]')

**Example 5.27** (2017)
**wenn es wäre** gut für den PC
('**if it were** good for the PC')

Besides concessiveness, a further explanation for parataxis following the *wenn* cluster clauses (*auch wenn*, *selbst wenn*, *sogar wenn*, and *und wenn*) is the observation by Breindl, Volodina, & Waßner (2014: 966–67) that they can function 'disintegrated' from the main clause, which may lead to the use of parataxis. Disintegration has been one of the main explanations for the occurrence of parataxis in previous scholarship, as we saw in Chapter 2, Section 2.2.4. In Example 5.28, *auch wenn* has a concessive function, and can be interpreted as slightly disintegrated from the main clause, as part of the stream of consciousness writing style used (which also makes it difficult to parse).

**Example 5.28** (2017)
[YOUTUBER] ich liebe dich danke wegen dem video es hat mir sehr geholfen! Du bist soo hübsch auch **wenn du bist** 100 videos drehst aber dann über Rollstühle haha spass werde ich dir soo lange zusehen bis ich

nicht mehr auf der welt exestiere ! Ich liebe dich [YOUTUBER]😍😘💄💝
('[YOUTUBER] I love you thank you for the video it really helped me! You are soo cute even **if you are** 100 and record videos but then with a wheelchair haha joke I will watch you soo long until I no longer exist in the world ! I love you [YOUTUBER]😍😘💄💝')

However, paratactic clauses are found in the *NottDeuYTSch* corpus following temporal uses of *wenn*, so the non-standard syntactic structure cannot be explained either by the concessive function of the conjunction nor by disintegration from the main clause. One possible reason is that these instances may be the result of non-native speaker usage, as is demonstrated in in Example 5.29. Here it seems as though Turkish syntactic structures have been used with German lexical items.

**Example 5.29** (2017)
Kannst du mal nach Österreich kommen, **wenn ich bin 18 Jahren** [werde ich] mein Führerschein bekommen. Fahre ich gleich nach Hamburg aber das hat noch seine zeit.
('Can you come to Austria some time, **when I am 18** [I will] get my driver's licence. I'll drive to Hamburg straight away but that will have its time')

Table 5.10 breaks down the syntax structure of Example 5.29, demonstrating the similarities to Turkish by aligning the phrases in the German and in a potential Turkish translation, with literal English, standard English, and standard German translations for comparison. By comparing the comment with the Turkish translation, we can see that the main clause in the comment aligns with the temporal adverbial clause in Turkish, followed by the noun phrase and verb in the subordinate clause, which are analogous to the position of the noun and verb phrases in Turkish.

Table 5.10: Potential multilingual emergent grammar structures in the *NottDeuYTSch* corpus

| | | | |
|---|---|---|---|
| **DE:** | wenn ich bin 18 Jahren | mein Führerschein | bekommen |
| **TR:** | 18 yaşıma geldiğimde/ 18 olduğumda | ehliyetimi | alacağım |
| **Lit. EN:** | When I turn 18 years/ When I am 18 | driver's licence-mine | (I will) receive |
| **St. DE:** | Wenn ich 18 Jahre alt bin, | werde ich meinen Führerschein | bekommen |
| **St. EN:** | When I am 18 years old | I will receive | my driver's license |

The alignment is not perfect: *bekommen* in the comment is presumably in the infinitive, whilst in Turkish it would be inflected for person and tense (*alacağım*). However, this could be a "grammatical reduction", characteristic of Turkish-German language contact (Wiese 2009: 786-787). It remains to be seen if similar constructions can occur with other conjunctions, or if this phenomenon is representative of other emergent developments potentially resulting from language contact, such as the use of V3 (verb-third) word order in German (see Bunk 2020).

This section has shown that the large size of the *NottDeuYTSch* corpus and quantitative approach taken by this study has facilitated the identification of emergent syntactic trends, both in terms of potential language contact and also the identification of parataxis following conjunctions that have not been considered in previous studies, partially due to their relative infrequency, such as *wenn*, *wenn* clusters (n = 12, 3.8 ipmc) and *solange* (n = 1). The corpus also enables the investigation of longitudinal trends to provide empirical data to address the claims in previous studies that the use of parataxis has been increasing since the 1980s (Günthner 1996: 323), to which I turn in Section 5.2.2.

### 5.2.2 Changes in parataxis over time

Over the period covered by the corpus (2008-2018), the proportion of clauses introduced by subordinating conjunctions containing parataxis has roughly doubled from 10 clauses per 1,000 to 20 clauses per 1,000 (Figure 5.3). This increase has largely been driven by the increase in paratactic *weil* clauses, but the use of paratactic *wenn* clauses has also increased since 2010 after an early fall (note that these figures exclude ambiguous clauses).

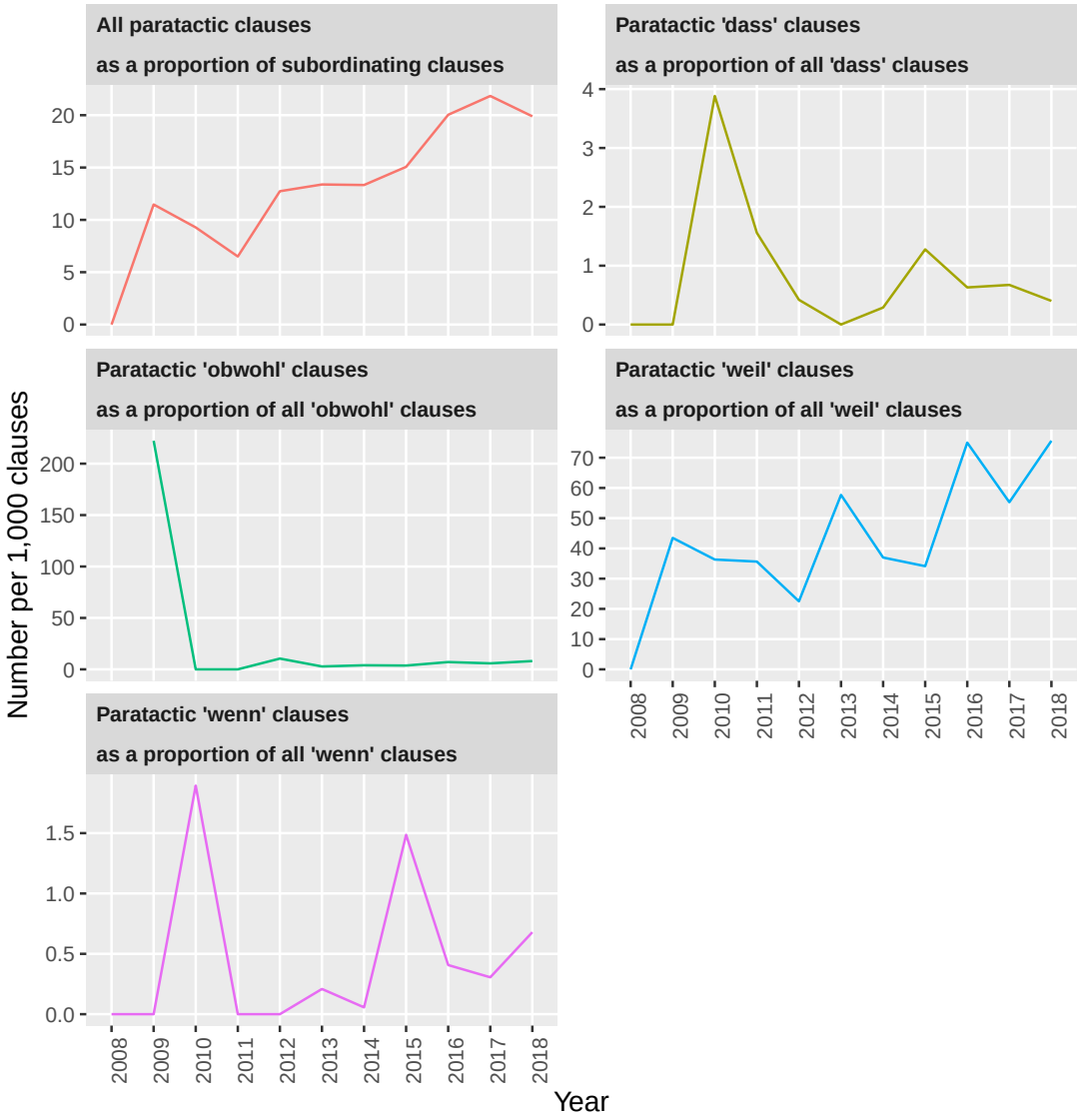



Figure 5.3: Frequency of paratactic clauses following the selected conjunctions in the *NottDeuYTSch* corpus over time

The frequency of paratactic *weil*-clauses increased considerably over the time-

frame of the corpus. In 2010, 43.4 *weil* clauses per 1,000 comments were paratactic; this had increased by 2018 to 75.6. The steady increase in frequency of paratactic *weil* clauses has two potential explanations: one general, one specific to the corpus, and both may hold true. First, *weil* may be undergoing gradual grammaticalization away from use as a subordinating conjunction to use as a coordinating conjunction and, potentially, a discourse marker. This possible explanation is explored more fully in Chapter 6, Section 6.2. Second, it may be that *YouTube* comments are, over time, increasingly using more features characteristic of spoken language, so that the frequency of paratactic *weil*-clauses is approaching that in the speech of young people in equivalent informal contexts, e.g. peer group conversations. To test this hypothesis would require substantial quantitative analysis with comparable spoken language data.

The other conjunctions do not exhibit the same rate of increase as *weil*. In fact, no trends are discernible, and the variations from year to year in the number of paratactic clauses are minor, despite the apparent 'spikes' in the proportions of paratactic clauses following *dass* and *wenn* in 2010 and the large drop in frequency following *obwohl* in 2009. These can be partly explained by the relatively low numbers of comments in the earlier years of the *NottDeuYTSch*, as the German *YouTube* industry was still a fledgling industry, meaning that lower raw numbers make smaller differences seem larger in Figure 5.3.[13] In March 2010, there was a redesign of the *YouTube* comment area to encourage more interaction between commenters, which could partly explain the increase in the number of comments since that time.[14] However, to more accurately measure any potential trends of the usage of paratactic structures with these subordinating clauses, either a (still) larger dataset is required, or the conjunctions would need to be monitored over a (still) longer period of time.

[13] In 2009, there were only nine *obwohl* clauses, two of them paratactic, compared to 191 *obwohl* clauses, two of which are paratactic in 2012, hence the apparent large graphic discrepancies in Figure 5.3.

[14] The newly designed comments section now placed the highest-rated comments at the top and introduced a tagging feature, where commenters could more easily respond to other comments by tagging the name of the commenter.

### 5.2.3 Comparing parataxis across corpora

No studies have yet explicitly compared the rates of parataxis of individual subordinating conjunctions between different corpora. Here, I compare the findings from the *NottDeuYTSch* corpus with data from three other corpora: the *DWDS Kernkorpus 21*, the *Dortmunder Chat-Korpus*, and the *DWDS WebXL Korpus*. As all three are tagged for parts of speech, it required relatively little manual processing to investigate them for paratactic clause structures. The following search function was initially used to count paratactic clauses following subordinating conjunctions:

```
$p=KOUS   #1   $p=/[ABCFIXNTP][A-Z]{2,6}/
$p=/V\wFIN/   $p=/([BCFIXNTP][A-Z]{2,6}|
A[DR][JT][AD]?)/
```

This search request is composed of four parts:

1. `$p=KOUS` matches a subordinating conjunction (this was then filtered to only include the subordinating conjunctions selected, as per Section 5.1.1)
2. `#1 $p=/[ABCFIXNTP][A-Z]{2,6}/` searches for 0 or 1 words followed by a part of speech that matches a particular set (that excludes verbs, for example, to reduce false positives)
3. `$p=/V\wFIN/` matches a finite verb
4. `$p=/([BCFIXNTP][A-Z]{2,6}|A[DR][JT][AD]?)/` matches another part of speech in a slightly different set (one that excludes verbs and prepositions to account for exbraciation, as investigated in Section 5.1.3.1). This also reduces the number of ambiguous cases captured, as it requires a part of speech to follow the verb

The results were then manually checked to remove other false positives. Table 5.11 presents rates of parataxis in the four corpora, broken down by the most frequent conjunctions in the *NottDeuYTSch* corpus.

Table 5.11: Comparison of the rates of parataxis across selected corpora

| Statistic | Corpus | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | DWDS Kernkorpus 21 | | DWDS WebXL Korpus | | Dortmunder Chat-Korpus | | NottDeuYTSch Corpus | |
| | Raw figure | Instances per million sentences | Raw figure | Instances per million sentences | Raw figure | Instances per million sentences | Raw figure | Instances per million sentences |
| dass | *0* | 0.00 | *281* | 1.41 | *0* | 0.00 | *41* | 10.42 |
| obwohl | *0* | 0.00 | *687* | 3.44 | *0* | 0.00 | *51* | 12.96 |
| solange | *0* | 0.00 | *11* | 0.06 | *0* | 0.00 | *1* | 0.25 |
| wenn | *1* | 1.14 | *85* | 0.43 | *0* | 0.00 | *58* | 16.26 |
| weil | *2* | 2.29 | *9 955* | 49.87 | *46* | 371.91 | *5 249* | 1 333.55 |
| other | *0* | 0.00 | *0* | 0.00 | *0* | 0.00 | *15* | 3.81 |
| **Total parataxis** | ***3*** | **3.43** | ***10 790*** | **55.20** | ***46*** | **371.91** | ***5 415*** | **1 377.25** |

[a] Figures for the DWDS WebXL Korpus based on stratified sample of 1,547 comments

The relative frequency of paratactic clauses as a proportion of all occurrences of each subordinating conjunction is similar across the four corpora. Overall, parataxis with *weil* is most frequent in all corpora, as expected, accounting for around 95% of all instances in the *DWDS WebXL Korpus*, *Dortmunder Chat-Korpus*, and *NottDeuYTSch* corpus. However, the data also confirm that parataxis following subordinating conjunctions is more frequent in colloquial language. Specifically, the rate of parataxis per million sentences in the four corpora forms a cline, where the rate of parataxis correlates with the degree of informality in a corpus (Figure 5.4). As we would expect in a corpus of formal written language, the *DWDS Kernkorpus 21* has the lowest rate of parataxis after subordinating conjunctions, followed by the *DWDS WebXL Korpus* and *Dortmunder Chat-Korpus*. Compared to the published webpage content of the *DWDS WebXL Korpus*, the higher rate in the *Dortmunder Chat-Korpus* is to be expected, as it is more chat-focused. The *NottDeuYTSch* corpus, as a collection of mostly informal communication written by young people, has the highest rate of parataxis of all, at almost four times the rate of the *Dortmunder Chat-Korpus* and over 450 times the rate in the *DWDS Kernkorpus 21*.

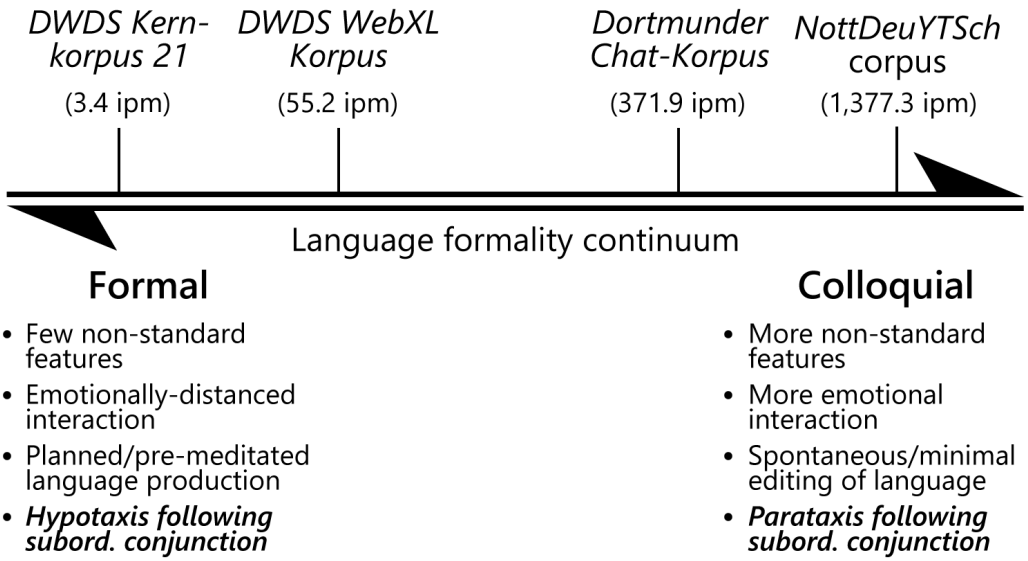



Figure 5.4: Continuum of syntactic characteristics based on language formality continuum with a visual indication of the formality of the selected corpora

The cline may well also reflect the proportion of younger contributors: while the age demographics of the other three corpora are not known, it is highly likely that the proportion of young people is dramatically lower in the other three cor-

pora. Regardless, the cline of proportions of parataxis in clauses introduced by subordinating conjunctions is quantitative evidence of a link between informality and grammatical change. This link is also demonstrated by the emergence of additional subordinating conjunctions to introduce paratactic clauses in the *NottDeuYTSch* corpus, namely, *solange*, *wenn*, and *wenn* clusters.

## 5.3 Future trends for parataxis in digital youth language

The emergence of additional subordinating conjunctions with paratactic functions, shown in Section 5.2.1, is no guarantee that they will continue to be used in such a way. However, the finding does demonstrate that large data sets can identify potential future trends to be monitored in subsequent investigations. Furthermore, the frequency of paratactic *weil* clauses in the *NottDeuYTSch* corpus has doubled between 2008 and 2018, as shown in Section 5.2.2, indicating the wider acceptance of paratactic *weil* clauses in digital youth language.

The comparative analyses of the corpora presented in Section 5.2.3 demonstrate that parataxis following a subordinating conjunction is not restricted to spoken language or written reproductions of direct speech, as posited in previous research. Admittedly, the adaptation of other spoken features into a written form has been seen as characteristic of digital writing, but, as argued in Chapter 2, classifying digital writing as mere written reproductions of speech does not accurately capture the rich and multimodal communicative styles in DMC. Indeed, the usage of conjunctions as coordinating conjunctions or discourse markers is far more frequent in digital youth language than in other sources of language, even other digital writing, and so offers further evidence for the characterisation of youth language as inherently linguistically creative and innovative (see Bahlo et al. 2019: 63). However, the relatively high frequency of use of subordinating conjunctions as coordinating conjunctions and discourse markers does not necessarily stem from some kind of language deficit, an accusation often still levelled at young people (see Chapter 2, Section 2.2). While there

are certainly commenters who unintentionally deviate from standard language, previous research has demonstrated that young people are often perfectly capable of both writing and speaking standard German, as well as non-standard varieties (e.g. Androutsopoulos & Busch 2021: 7).

Having identified the extent of the developments in the frequency of paratactic clauses occurring in clauses introduced by subordinating conjunctions, Chapter 6 turns to the challenge of explaining how and why they occur. Using data from the *NottDeuYTSch* corpus, the chapter addresses previous work that has studied the grammatical processes purportedly leading to the use of paratactic *weil* clauses, in order to assess the applicability of previously proposed explanations to digital youth language.

# 6

# From subordinating conjunctions to discourse markers? Investigating grammaticalization in digital youth language

Following on from the initial overview of parataxis in clauses introduced by subordinating conjunctions in Chapter 5, this chapter examines the syntactic data from the *NottDeuYTSch* corpus in more detail. It tests explanations proposed in previous research for the occurrence of parataxis, and proposes new semantic hypotheses based on the data in the *NottDeuYTSch* corpus.

Section 6.1 presents a data-driven examination of the *NottDeuYTSch* corpus to investigate whether the first field constituent in clauses introduced by a subordinating conjunction significantly influences the likelihood of parataxis. The analysis reveals that the high frequency of particular parts of speech in paratactic clauses, such as conjunctive adverbs (e.g. *dann* and *eigentlich*), is indicative of a change in the function of the subordinating conjunction introducing the clause: namely, it has undergone grammaticalization such that it functions as a discourse marker in these cases.

Section 6.2 then investigates in greather depth how subordinating conjunctions become grammaticalized allowing them to function as discourse markers and coordinating conjunctions. The section first tests the existing hypotheses that potentially explain this process, as well as existing models of grammaticalization, using the data in the *NottDeuYTSch* corpus to establish whether they account for the grammatical changes described in both Chapter 5 and Section 6.1. In particular, the section develops a new grammatical model, building on previous studies, to describe the grammatical processes and developments in digital youth language.

## 6.1 The role of the first field in paratactic clauses

While previous studies of parataxis in clauses introduced by subordinating conjunctions have tended to focus on the relationship of the subordinating conjunction to the main clause, this section examines the occurrence of parataxis in relation to the lexical item in the first field of the clause. One potential reason for the paucity of such studies in existing scholarship is the small size of the data sets used and the mostly qualitative approach. The relatively large quantity of data from the *NottDeuYTSch* corpus (i.e. over 33m words) enables the use of corpus linguistic quantitative methodology to detect several emergent patterns within this syntactic phenomenon not been previously identified in literature. To carry out the analyses in this section, all hypotactic, paratactic, and ambiguous clauses following the 30 subordinating conjunctions identified in Chapter 5 were extracted, and the words and phrases in the clauses were categorised according to their position in the clause. This resulted in 325,075 hypotactic, paratactic, and ambiguous clauses in total to be analysed, of which 5,404 exhibited parataxis, as shown in Chapter 5, Section 5.2.1.

One study that examined the content of the subordinate clause is Kempen & Harbusch (2016: 3), who found that a personal or indefinite pronoun, such as *ich* or *man*, occurred following *weil* in "only a minority of the V2 [verb-second] clauses", in their data from the *PhonDat-VERBMOBIL* corpus of spoken language (Hess, Kohler, & Tillmann 1995: 863). However, the data in the

*NottDeuYTSch* corpus demonstrates a different trend, with pronouns occurring in the first field in 63% (n = 3,417 of 5,404) of paratactic clauses with a subordinating conjunction (i.e. *dass*, *obwohl*, *solange*, *weil*, *wenn*, and *wenn* clusters). Reviewing the qualitative examples of parataxis provided in Reis (2013), Antomo & Steinbach (2010), and Freywald (2016), we find that 53% (n = 38) contain personal or indefinite pronouns in the first field, similar to the findings from the *NottDeuYTSch* corpus. No firm conclusions can be drawn from these limited data, especially as the three studies cited did not specify their sources, but the finding from the *NottDeuYTSch* corpus appears to be in line with other available data on such structures.

Of all the 325,075 clauses introduced by the 30 selected conjunctions in the *NottDeuYTSch* corpus (regardless of syntax structure), 65% (n = 212,837) contained a pronoun in the first field compared to 63% of paratactic clauses. This indicates that, despite their high overall frequency, pronouns occur roughly as frequently in paratactic clauses in all clauses in the *NottDeuYTSch* corpus, and also shows that the use of a pronoun in the first field does not have a significant effect on the occurrence of paratactic syntax structure. However, the occurrence of other lexical items in the first field could influence the structure of the clause. To investigate this, the lexical items with the largest disparity between their frequencies in paratactic clauses and all clauses were statistically analysed. This analysis was restricted to *weil* clauses to ensure consistency of results due to the high frequency of paratactic *weil* structures in the *NottDeuYTSch* corpus (97%). The analysis revealed that the majority of first field constituents with the largest disparities between their occurrence in paratactic and all clauses are what are referred to as predicative first field constituents, as shown in Table 6.1, i.e. lexical items that are not the grammatical subject or object in the clause, such as *früher*, *irgendwann*, *dann*, and *eigentlich*.[1] In Example 6.1, *dann* occurs in the first field, following the conjunction *weil*, while the subject, [YOUTUBER2], is in the midfield of the clause.

[1] *Eigentlich* often occurs as a particle in the *NottDeuYTsch* corpus too, but where it occurs in the first field of a paratactic clause following a subordinating conjunction, it functions as an adverb.

**Example 6.1** (2015)
Echt coole Idee kann bitte [YOUTUBER] mitspielen **weil dann kann** [YOUTUBER2] wieder sein Haus sprengen
('Really cool idea can [YOUTUBER] play with you **because then** [YOUTUBER2] **can** blow up his house again')

Further analysis of paratactic *weil* clauses reveals several potential explanations for the high relative frequency of predicative first field constituents, that, I shall argue, goes beyond simply topicalisation (*Topikalisierung*), i.e. the placing of a lexical item that is not the subject, such as complement, adverbial phrase or non-finite verb in the first field (Leibniz-Institut für Deutsche Sprache 2018c). To do so, I analyse the predicative first field constituents that can be classified as conjunctive adverbs, an adverb that semantically links to the previous sentence or clause, as explained in Chapter 5, Section 5.1.1. Conjunctive adverbs (as per the list in Breindl, Volodina, & Waßner (2014: 20)) that appear in the first field of paratactic clauses include *bald*, *damit*, *dann* (the adverb with the highest number of occurrences, 183), *davor*, *eigentlich*, *einmal*, *erst/erstens*, *mittlerweile*, *somit*, *sonst*, *soweit*, *vorher*, and *zuerst*. I was able to find one comparable occurrence of a conjunctive adverb in an example paratactic clause in Reis (2013: 247) (see Example 6.2).

**Example 6.2** (Reis (2013: 247))
Paul hat den Job bekommen, **weil sonst** käme er nicht immer im Anzug daher.
('Paul got the job, **because otherwise** he would not always come from there in a suit')

Table 6.1: First field constituents with the largest disparity between their frequencies in paratactic *weil* clauses and all *weil* clauses in the *NottDeuYTSch* corpus

| | Paratactic 'weil' clauses | | All 'weil' clauses | | Observed versus expected frequencies of parataxis | |
|---|---|---|---|---|---|---|
| **First field constituent** | **Raw** | **Instances per 1,000 clauses** | **Raw** | **Instances per 1,000 clauses** | **Increase above expected freq (%)** | **Significance** |
| früher | *5* | 0.97 | *6* | 0.03 | 3 758.6 | p<0.001 |
| irgendwann | *7* | 1.35 | *12* | 0.05 | 2 601.0 | p<0.001 |
| sowas | *24* | 4.64 | *42* | 0.18 | 2 545.9 | p<0.001 |
| dann | *183* | 35.41 | *341* | 1.43 | 2 384.9 | p<0.001 |
| YOUTUBER | *8* | 1.55 | *17* | 0.07 | 2 079.0 | p<0.001 |
| geld | *7* | 1.35 | *15* | 0.06 | 2 060.8 | p<0.001 |
| eigentlich | *17* | 3.29 | *37* | 0.15 | 2 027.5 | p<0.001 |
| was | *19* | 3.68 | *42* | 0.18 | 1 994.7 | p<0.001 |
| damit | *13* | 2.52 | *30* | 0.13 | 1 906.5 | p<0.001 |
| sonst | *70* | 13.54 | *168* | 0.70 | 1 829.3 | p<0.001 |

In addition to topicalisation as a possible trigger for how conjunctive adverbs function in the first field of a paratactic clause following a subordinating conjunction, I present three additional interpretations that could explain this innovative grammar construction. The four potential functions, listed below, are explained in greater detail below using data from the *NottDeuYTSch* corpus:

1. **Topicalisation:** conjunctive adverbs simply function in the first field like other adverbs and are "topicalised" and emphasised.
2. **Conjunction clusters:** conjunctive adverbs form a novel conjunction cluster (such as *weil dann* or *weil sonst*) with the preceding conjunction.
3. **Multiple fronting:** Conjunctive adverbs exhibit "multiple fronting" with other lexical items, such as noun phrases, where several constituents occur in the first field (e.g. the indirect and direct objects of the sentence in *Kindern Bonbons sollte man nicht geben*, see Müller 2003: 2-3, 2019: 395-397; Bildhauer & Cook 2010: 2-3).
4. **Discourse marker:** When a conjunctive adverb occupies the first field, the preceding conjunction in fact functions as a discourse marker.

**1. Topicalisation.** It is possible that the trigger here is topicalisation. However, 20.4% (n = 356) of 1,745 *weil* clauses with a predicative first field exhibit paratactic word order, more than four times the 4.37% proportion (n = 3,591) of 82,174 *weil* clauses with the subject in the first field which exhibit paratactic word order. This difference is statistically significant, $\chi^2$ (1, n = 83,919) = 979.7482, $p < 0.001$. The large discrepancy suggests that beyond simple topicalisation, there is an interaction between word order after these conjunctions and the elements of the first field.

**2. Conjunction clusters.** A second possibility to explain the high frequency of predicative first field constituents is the occurrence of conjunction clusters. Examples 6.3 and 6.4, *weil dann* and *weil sonst* demonstrate potential novel clustering of conjunction and conjunctive adverb that function as a single lexical item (e.g. *als ob*, *und zwar*, or *nur dass*), in this case as a conjunctive adverb, because the verb immediately follows the cluster. However, in Examples 6.5

and 6.6, *weil* and *dann* would form a cluster located in the left margin. They are followed by a first field lexical item, which indicates that the clusters can function as a subordinating conjunction. Semantically, these clusters express a conditional (*weil dann*) or negative conditional (*weil sonst*) semantic relationship with the main clause.

**Example 6.3** (2015)
Echt coole Idee kann bitte [YOUTUBER] mitspielen **weil dann kann** [YOUTUBER2] wieder sein Haus sprengen
('Really cool idea can [YOUTUBER] play with you **because then** [YOUTUBER2] **can** blow up his house again')

**Example 6.4** (2015)
Ich hoffe sehr das dass nur erfunden ist **weil sonst wäre** das echt krank…
('I seriously hope that that is only made up **because otherwise** that **would be** really sick')

**Example 6.5** (2017)
**Weil dann** meine schwester will eine ps4 haben
('**Because then** my sister wants to have a ps4')

**Example 6.6** (2017)
Eine frage an die hater warun hatet ihr [YOUTUBER] später wenn ihr älter seid oder ob ihr schon älter seid keine Ahnung werdet ihr nicht mehr [YOUTUBER] haten **weil dann arm seid ihr [IDIOTEN]**
('A question for the haters why are you hating on [YOUTUBER] later when you are older or if you are already older no idea you will not hate on [YOUTUBER] any more **because then you [IDIOTS] will be poor**')

**3. Multiple fronting.** Alternatively, Examples 6.5 and 6.6 could be analysed as cases of multiple fronting, where "under certain conditions, several constituents can be found [in the first field]" (Müller 2003: 2-3, cf. 2019: 395-397; Bildhauer & Cook 2010: 2-3), in this case "Subject and Adverb" fronting (Müller 2003: 3). In Example 6.5 we find *dann* as the adverb, and *meine Schwester* the subject.

Example 6.6 presents a combination of an adverb, *dann*, and the predicative adjective *arm*, a kind of contribution which is not explicitly described by Müller.

A stronger candidate for clustering is exhibited in comments with *weil wegen*, as demonstrated in Examples 6.7 and 6.8 (translated as 'because cos of'), which appears 18 times in the *NottDeuYTSch* corpus (accounting for 0.34% of all paratactic *weil* clauses). Unlike the other candidates for a novel cluster (*weil dann*, *weil sonst*), there is considerable semantic overlap between the two lexical items.

**Example 6.7** (2012)
Ich möchte gewinnen **weil wegen** is so ;D
('I would like to win **because cos of** it is like that ;D')

**Example 6.8** (2012)
Ich will gewinnen ,**weil wegen** Wurst
('I want to win ,**because cos of** sausage')

However, further research is needed to establish if this phenomenon is indeed multiple fronting, especially since existing research on the topic has focused on the occurrence of the phenomenon in main clauses, not subordinate clauses.

**4. Discourse markers.** A fourth possibility is that, in Examples 6.2 to 6.6, *weil* can be interpreted as structuring the flow of the comment, rather than introducing a causal connection between the clauses or syntactically linking to the previous clause; that function is performed by the conjunctive adverb. Under this interpretation, *weil* functions as a discourse marker, rather than as a paratactic conjunction (or as part of a conjunction cluster). As a discourse marker, it can either introduce additional information, a narrative sequence, a change of topic, or structure a conversation (see Chapter 2, Section 2.2.4).

A further explanation to support the interpretation of *weil* in these cases as a discourse marker is the function of *wegen* in Example 6.8, which fulfils the causal connection between the two clauses, not *weil*. Here, *weil* is inserted after the full verb in the right bracket, *gewinnen*, i.e. the verb that carries meaning, and before what can be seen as an exbraciated prepositional phrase, *wegen baum*, which

occupies the final field position (despite the use of a comma preceding *weil*). A case has often been made for *weil* and other subordinating conjunctions functioning as discourse markers, as well as in the role of coordinating conjunctions, i.e. they have undergone or are undergoing grammaticalization (e.g. Günthner 1993; Gohl & Günthner 1999; Freywald 2016; Frey & Masiero 2018), and this hypothesis warrants further consideration. In Section 6.2, I therefore compare the findings of existing research on the functions of subordinating conjunctions from the *NottDeuYTSch* corpus.

## 6.2 Grammaticalization of subordinating conjunctions to coordinating conjunctions and discourse markers

This section examines the potential grammatical processes that can explain the increase in the frequency and variety of subordinating conjunctions introducing paratactic clauses. In addition to grammaticalization, the section tests other explanations presented in previous studies on the use of parataxis rather than hypotaxis in clauses introduced by conjunctions to the data of the *NottDeuYTSch* corpus. However, grammaticalization best describes the grammatical developments identified in digital youth language in this chapter. Therefore, the section examines how grammaticalized forms of subordinating conjunctions can be identified in Digital Writing through the analysis of orthographic features and contextual clues, in particular, considering in detail the distinguishing features between the use of the same lexical item either as a coordinating conjunction or as a discourse marker. Having identified the cases of grammaticalization of subordinating conjunctions in the *NottDeuYTSch* corpus, the section applies these findings to the model of grammaticalization presented by Gohl & Günthner (1999), suggesting a subsequent refinement of the model based on the emergent grammatical forms in the corpus.

### 6.2.1 Applying existing hypotheses of parataxis to the *NottDeuYTSch* corpus

In Chapter 5 four overlapping hypotheses for parataxis were investigated:

1. Parataxis may occur in clauses with illocutionary autonomy from the main clause (Reis 2013; Freywald 2016).
2. Parataxis may occur in clauses that are syntactically disintegrated from the main clause (Antomo & Steinbach 2010).
3. Parataxis may occur in *weil* clauses where *weil* is used analogous to *denn* (Reis 2013).
4. Parataxis may occur because the subordinating conjunction has undergone grammaticalization which enables it to function as a coordinating conjunction or a discourse marker (Freywald 2016).

The four hypotheses are addressed in turn below, drawing on the data from the *NottDeuYTSch* corpus.

Example 6.9 shows a comment made in a response to a listicle-style video, in which a popular viral video dance is performed 17 times using different styles, hence the 17 points in the comment. The commenter suggests that a subsequent video could be made in which each of the dance styles is influenced by the characters from 'School of Dragons', a video game spin-off of the 'How to Train your Dragon' film franchise. The example contains eleven hypotactic and three paratactic *weil* clauses in the same comment, yet each of these *weil* clauses seems to perform the same function. Applying the concept of illocutionary autonomy in Reis (2013) and Freywald (2016) does not produce satisfactory result. None of the fourteen *weil* clauses contain any independent illocutionary acts: they are all causal reasonings for the choice of dragon (even the three clauses that are not introduced by a conjunction), yet there is a mix of paratactic and hypotactic word order.

**Example 6.9** (2017)

MEGA!!! Hmmmmmmm (Nur wer School of Dragons [a computer game] kennt (; mit Begründung)

1. Riesenhafter Albtraum *weil ich was Klassisches nahm*
2. Armorwing *weil sie eine ähnliche aura haben*
3. Rumpelhorn Ich mag weder den Drachen noch Dabs [dance move]„
4. Tödlicher Nadder *weil das irgendwie für mich gepasst hat*
5. Gleidgesang *weil ich Psy [music artist] gefeiert hab*
6. Schneller Stachel, ich fand Mal der sollte was cooles bekommen
7. Sand Geist *weil das für mich wieder Mal gepasst hat*
8. Wechsel Flügler *weil auch so ein Drache Mal sanft sein kann*
9. Brüllender Tod *weil, es der Drache in Person ist*
10. Nacht Terror (Farbe: weiß) Die sind sehr klein und knuddelig perfekt für so eine
11. Silber Phantom (neuer Drache in SoD) *weil Dieser Drache für ihn Irgendwie Passt*
12. Schnappende Falle+Rauch Atem *weil die Schnappende Falle nur 4 Köpfe hat* brauchte ich noch was kleines für die Mitte
13. Skrill+Schreckliche Schrecken **weil er hat diesen Drachen so verdient** und die Schrecken als backround Sänger, ihr wisst welche Stelle ich meine
14. Woll Geheul **weil der Drache hat solche Haare wie er**
15. Holzklau **Weil Ich liebe Skrillex** [music artist] und ich Liebe den Holzklau
16. Todsinger *weil Siri [digital assistant from Apple] meine Top 4 von den allen war* und der Todsinger auch meine Top 4 Drachen ist
17. Flutsegler *weil [YOUTUBER] auch Mal was halbweg nices Verdient hat*

Hoffe es hat euch gefallen <3

('AMAZING!!! Hmmmmmmm (only if you know School of Dragons (; with reason)

1. Monstrous Nightmare *because I'll take something Classic*
2. Armorwing *because they have a similar aura*

3. Rumblehorn I don't like either the dragon or dabs„
4. Deadly Nadder *because that kind of suited me*
5. Slither Song *because I enjoyed Psy*
6. Speed Stingers I thought Maybe he should get something cool
7. Sand Wraith *because that suited me Again*
8. Changewing *because even such a dragon can be gentle*
9. Screaming Death *because, it is the personification of a dragon*
10. Night Terror (colour: white) They are very small and cuddly perfect for such a thing
11. Silver Phantom (new dragon in SoD) *because this dragon suits him somehow*
12. Snaptrapper+Smokebreath *because the Snaptrapper only has 4 heads* I needed something small for the middle
13. Skrill+Terrible Terror **because he really earned** these dragons and has the Terror as a background singer, you know what part I mean
14. Wooly Howl **because the dragon has** hair like him
15. Timberjack **Because I love** Skrillex and I Love the Timberjack 16. Death Song *because Siri was my top 4 of them all* and the Death Song is also my top 4 dragon
17. Tideglider *because [YOUTUBER] has also earnt himself something half nice*

Hope you enjoyed it <3')

The mix of parataxis and hypotaxis in *weil* clauses in Example 6.9 also means that the guidelines in Antomo & Steinbach (2010) to justify the choice of syntax structure based on the integration of the *weil* clause with the main clause cannot be consistently applied. For example, in point 14 (*weil der Drache hat solche Haare wie er*), the commenter selects dragons based on the apparent physical similarities with the *YouTuber*, which, I would argue, is a strong causal link. Such a structure, according to Antomo & Steinbach (2010), should be hypotactic due to the syntactic integration with the antecedent main clause. However, the *weil* clause in this instance is paratactic. Furthermore, point 5 contains a hypotactic structure, yet the causality between the two clauses (*weil ich Psy gefeiert hab*)

is less strongly linked to the choice of dragon than in point 14, as the reader must infer the link (or is assumed to know the link) between the dragon and *Psy*, the Korean musician. Possibly the intended but tenuous link in point 5 is that the name of the dragon is also musically-related (Slither Song/*Gleidgesang*). There is a similarly weak causal link in point 15 where the commenter equates liking the musician, *Skrillex*, to liking a dragon (*Weil Ich liebe Skrillex und ich Liebe den Holzklau*), perhaps in this case associating the plaid shirts worn by the artist with the lumberjack-theme of the dragon. However, in this case, the commenter does use a paratactic *weil* clause.

There are further examples of clauses that are almost lexically identical yet with differing syntax structures, such as in points 13 and 17, which employ the same verb and tense (*verdienen*, in the perfect tense). Another instance of very similar instances with differing word orders is given in Example 6.10, where the only difference is the subject (*weil [YOUTUBER1] einfach geil ist* and *weil es ist einfach geil*).

> **Example 6.10** (2012)
> *weil [YOUTUBER1] einfach geil ist* !!! ich würde mich sehr sehr freuen wenn ich gewinnen würde… **weil es ist einfach geil** von so ner coolen person ein autogramm und kino-gutscheine zu bekommen. bin ein großer fan von dir und euch allen anderen von [YOUTUBER2] auch.
> ('*because [YOUTUBER1] is simply cool* !!! I would be so so happy if I won… **because it is** simply cool to get an autograph and cinema vouchers from such a cool person. I'm a big fan of you and all of you others from [YOUTUBER2] as well.')

Again there are no clear differences in syntactic integration or illocutionary autonomy between the two pairs of clauses in points 13 and 17 of Example 6.9 and Example 6.10, either on a lexical or orthographic level. This suggests that the hypotheses of syntactic integration and illocutionary autonomy are not applicable to the data in the *NottDeuYTSch* corpus.

Furthermore, the third hypothesis mentioned above, i.e. that paratactic *weil*

clauses can occur if *weil* could be substituted for *denn*, is also not consistently applicable to the corpus. A case where it cannot apply is demonstrated in Example 6.11,[2] where a paratactic *weil* clause appears antecedent to the main clause, which is incompatible with the syntactic function of *denn* (see Reis 2013; Scheffler 2005; Pasch et al. 2003).

**Example 6.11** (2016)
**weil es kamen viele neue updates und neue dlc packs in ark** *bitte spiel es mal wieder mit [YOUTUBER]*
('**Because lots of new updates and new dlc packs in ark have come out** *please play it again some time with [YOUTUBER]*')

This leaves the fourth hypothesis, that some subordinating conjunctions have undergone grammaticalization where they function as coordinating conjunctions and discourse markers. This seems to account for the data in the *NottDeuYTSch* corpus, as will be argued in Section 6.2.2 below. However, it is not clear why the other previous hypotheses are (no longer) applicable to the data. While the previous hypotheses may have been valid for the language at the time, the high frequency of paratactic structures in clauses introduced by subordinating conjunctions in the corpus suggests that the phenomenon has been incorporated into digital youth language for use in less restricted circumstances. This, in turn, has caused overall usage of the phenomenon to increase, as well as the number of subordinating conjunctions that are used to introduce paratactic clauses.

### 6.2.2 Distinguishing between coordinating conjunctions and discourse markers in the *NottDeuYTSch* corpus

It has also been suggested in previous scholarship that some subordinating conjunctions, including *weil* and *obwohl*, may function as discourse markers, in addition to possible grammaticalization where they function as coordinating conjunctions. Spoken language has overwhelmingly been the source of data

[2]The *weil* antecedent is in bold and the main clause is in italics.

for such analyses (with some exceptions in Freywald (2016), who included a few examples from internet forums). Examples 6.12 to 6.14 are taken from previous work. In each case, the subordinating conjunctions are explicitly described as discourse markers (relevant clauses in bold):

**Example 6.12** (Günthner (1993: 37))
also **weil - man kann es ja wissenschaftlich untersuchen**.
('so **because - one can investigate it scientifically of course**')

**Example 6.13** (Freywald (2016: 338))
Ja, leider hab ich morgen Schule Hm.. aber **obwohl, geh ich halt ohne Schlaf hin**, hab morgen eh nur 8 Stunden
('Yes unfortunately I have school in the moring hmm.. **although, I'll go in without sleeping**, I've only got 8 hours after all')

**Example 6.14** (Frey and Masiero (2018: 73))
A: Ich bin unzufrieden, wie es gelaufen ist.
B: Warum?
A: **Weil, die Leute haben gelangweilt dreingeschaut**. Es wurden wenige Exponate gekauft. Viele Besucher sind früh gegangen.
('A: I am unhappy at how it went
B: Why?
A: **Because, the people looked in bored**. Not many exhibits were bought. Many visitors went early.')

In spoken language, the discourse marker function of subordinating conjunctions has been identified through pauses or intonation patterns in speech (see Gohl & Günthner 1999: 47-48). When transcribing spoken language, commas or dashes either following or surrounding the discourse marker have been used, but in spontaneously written language, particularly Digital Writing, discourse markers are less readily identified. They may sometimes be indicated by punctuation following the conjunction (Schäfer & Sayatz 2016: 228), but for the *NottDeuYTSch* corpus this cannot be relied upon, as the language is characterised by a lack of standard punctuation, i.e. either it is not used at all (see

Brommer 2007; Herring 2008) or is used creatively (see Chapter 7). Example 6.15 from the *NottDeuYTSch* corpus contains two *weil* clauses, the first with paratactic structure (in bold) and the second with hypotactic structure (in italics). As we shall see, establishing whether the clause following the first *weil* would be classified as a coordinating conjunction or as a discourse marker proves difficult.

> **Example 6.15** (2012)
> Ich will (möchte) gewinnen **weil… Ähm… Ich habe noch die Ja Apfelsxhorle probiert** und *weil ich ma wieder mit meinen Freunden was unternehmen will*. Danke Danke
> ('I want (would like) to win **because… erm… I tried the Ja [own-brand from the Rewe supermarket chain] apple spritzer** and *because I want to do something again sometime with friends*. Thank you thank you')

The first *weil* can be interpreted as a coordinating conjunction following Antomo & Steinbach (2010) and Reis (2013). Here, the reference to the apple spritzer in the paratactic *weil* clause (in bold) is an attempt by the commenter to build rapport with the *YouTuber*. It is less directly related to providing a reason as to why the commenter should win the competition, compared to the hypotactic *weil* clause later in the same comment (in italics). It is, therefore, less syntactically dependent on the main clause (or less integrated, in the terminology of Antomo & Steinbach 2010), in comparison to the hypotactic *weil* clause, and can be interpreted as functioning as a coordinating conjunction. Under this reading, the conceptually oral use of ellipses and the filler word, *ähm*, are an expression of hesitation and indecision. However, it is likelier that these conceptually oral features support the interpretation of the first *weil* as a discourse marker. Rather than the punctuation and *ähm* simply providing metacommunicative 'flavour', their use lends weight to the argument that the first and second *weil* clauses have different linguistic functions. The punctuation and filler word induce the spatial separation of *weil* from the following clause, *ich habe noch die Ja Apfelsxhorle probiert*. This therefore indicates that the *weil* is introducing a conversational aside, functioning as a discourse marker, before the

commenter comes, in the hypotactic causal clause, to the real reason why they should win the competition.

The interpretation of the paratactic *weil* in Example 6.15 as either a coordinating conjunction or discourse marker is possible, and a similar difficulty of distinguishing between a coordinating conjunction and a discourse marker is demonstrated in the nascent research on paratactic *dass* clauses (of which the *NottDeuYTSch* corpus offers 41 instances). Freywald (2016: 333–34) provided several examples dating back to 2003 of *dass*-V2 in spoken language. However, Freywald (2016: 351–52) argued that when *dass* is followed by a paratactic structure, it functions as a "verbalised colon": in effect, one could remove *dass* from the clause and the clause would function as a grammatically standard "non-introduced subordinate clause", as demonstrated in 6.16, adapted from Freywald (2016), and 6.17 from the *NottDeuYTSch* corpus.

**Example 6.16** (adapted from Freywald (2016))

1. Paratactic *dass*:
Ich hab gelesen, **dass in Sizilien gibt's welche**, die sind 'n paar hundert Jahre alt.
('I read **that in Sicily there are some**, they are a few hundred years old.')

2. "nicht-eingeleiteter Nebensatz":
Ich hab gelesen, **in Sizilien gibt's welche**, die sind 'n paar hundert Jahre alt.
('I read **in Sicily there are some**, they are a few hundred years old.')

**Example 6.17** (2014)

1. Paratactic *dass* (original):
Ich finde, **dass du hast voll schöne Augen**;):*
('I think **that you have such beautiful eyes**;):*')

2. "nicht-eingeleiteter Nebensatz" (adapted):
Ich finde, **du hast voll schöne Augen**;):*
('I think **you have such beautiful eyes**;):*')

Freywald (2016: 333–34) claimed that the *dass* in such clauses "does not appear

as a conjunction […] but fulfils the function of a speech act, namely an assertion marker", and that "the V2 sentences linked with *dass* always represent a statement claimed to be true". Freywald's use of the term 'assertion marker' arguably implies that *dass*, when followed by main clause structure, functions as a discourse marker rather than as a coordinating conjunction. However, Freywald does not seem to allow for *dass* as a discourse marker in a clause where there is no statement claimed to be true, as in Example 6.18. In this example, the *dass* is embedded within a string of questions, where the speaker is seeking confirmation of a statement, rather than asserting its truth, almost being used in place of *ob*, the conditional subordinating conjunction. In this case, the *dass* does indeed seem to function as a coordinating conjunction, rather than as a discourse marker. Furthermore, Auer & Günthner (2003: 6–7) presented the argument that if another element can be inserted between the conjunction, which occupies the pre-first field, and the following main clause, and if such an element can itself fulfil the function of a discourse marker, the conjunction also therefore functions as a discourse marker. In the case of paratactic *dass* clauses, such as Example 6.18, other elements cannot be inserted between *dass* and the first field without substantially altering the structure of the sentence, therefore reinforcing the interpretation that *dass* can indeed function as a coordinating conjunction, not just as a discourse marker, when introducing paratactic clauses.

**Example 6.18** (2016)

[+USER] in dem viedeo waren sexuelle Inhalte ein Bild hat man die Brüste von einer Frau gesehen sie hatte ein Koch schürtze an falls das jemand ließt kann mir jemand beantworten **dass [YOUTUBER] hat gesagt** sie hat es getan wieso sie haben doch erst dannach geredet also ist seine Freundin ein Arschloch?

('[+USER] there was sexual content in the video a picture one saw a woman's breasts she had an apron on if anybody reads this can someone answer me **that [YOUTUBER] said** she did it how come they spoke first only afterwards so is his girlfriend an arsehole?')

Further debate on the classification of the function of subordinating conjunc-

tions that introduce paratactic clauses concerns situations where *weil* is followed by a W-question (*wann*, *warum*, *was*, *wer*, *wie*, *wo*) (see Examples 6.19 and 6.20), as well as *weil* followed by verb-first questions, as in Examples 6.21 and Example 6.22 (see Auer & Günthner 2003; Reis 2013; Freywald 2016). In the *NottDeuYTSch* corpus, there are also instances of W-questions introduced by *obwohl*, as Example 6.23 demonstrates.

**Example 6.19** ('W'-question: Reis (2013: 224))
**weil wo ist** es schöner im Juni als dort?
('**because where is** it more beautiful in June than there?')

**Example 6.20** ('W'-question: 2016)
Finde ich nicht **weil warum hat er** geweint
('I don't think so **because why did he** cry')

**Example 6.21** (verb-first question: Reis (2013: 225))
**weil würde er** sonst dauernd verreisen können?
('**because would he** otherwise be able to constantly travel?')

**Example 6.22** (verb-first question: 2017)
[YOUTUBER] (transcribed): dann schreibt ihr unterm Video welches Kissen ihr haben wollt und warum
[COMMENTER]: von dir... **weil ist es** nicht offensichtlich :3?
('[YOUTUBER]: then write under the video which cushion you want and why'
[COMMENTER]: [I want the cushion] Of you... **because is it** not obvious :3?')

**Example 6.23** ('W'-question following obwohl: 2013)
das schreit nach nen neuen schtempel schlechter als warrock [computer game] **obwohl was kann** schlechter als warrock sein!
('this sounds an awful lot like for a new stamp worse than warrock although what can be worse than warrock')

Reis considers all of these to be clear instances of the same *weil*-sentence type

as paratactic *weil* clauses (Reis 2013: 224), i.e. *weil* functions here as a coordinating conjunction. This differs from the interpretations of Auer & Günthner (2003) and Freywald (2016), who both considered *weil* and *obwohl* in these cases to function as discourse markers. Auer & Günthner (2003: 6) argued that *weil* and *obwohl* in such instances link clauses "in a more indirect way", i.e. epistemtically and with regards to the structure of the speech acts. Similarly, Freywald (2016: 329–30) implies that *weil* organises the "speech acts", which would categorise these instances as discourse markers. Therefore, the function of *weil* and *obwohl* in Examples 6.19 to 6.23 is, I contend, to provide additional information and to structure the information, two of the main functions of a discourse marker, rather than to coordinate the two clauses.

While the function of the conjunctions in Examples 6.19 to 6.23 may be identifiable using the framework of Auer & Günthner (2003) and Freywald (2016), the function is not always as apparent in other cases. Ágel (2016: 93) considers that paratactic 'subordinate' clauses, standard subordinating conjunctions, and discourse markers have some shared metapragmatic functions, which makes it difficult to develop a holistic framework to apply to written language in order to identify the function of a conjunction. On the one hand, the subordinate clause links to the main clause. For example, a clause introduced by *weil*, may substantiate the facts. Yet, at the same time, the conjunction also has discourse-organising functions like those that are attributed to discourse markers. For example, the *weil* in a *weil* clause could refer to a topic of shared knowledge between participants (see Blühdorn, Foolen, & Loureda 2017: 11). Without further evidence, the hypotheses of both Agel and Auer & Günthner seem to be valid. In Example 6.24, one could insert another discourse marker, such as *ja*, *also*, or *ich weiß nicht* (e.g. "Ich will das [YOUTUBER]-kissen **weil**, *ja*, ich könnte dir jetzt so eine richtig bescheuerte und unnötige Story schreiben"), which would suggest that the *weil* is a discourse marker according to Auer & Günthner (2003). However, it links semantically to the main clause (*von dir*) and organises the discourse, in accord with the definition of a coordinating conjunction in Ágel (2016).

**Example 6.24** (*weil*: 2017)
Ich will das [YOUTUBER]-kissen **weil**, ich könnte dir jetzt so eine richtig bescheuerte und unnötige Story schreiben so von wegen ja sonst könnte ich es mir nicht leisten und ja ich hab so ein trauriges leben tralala hopsasa […]
('I want the [YOUTUBER] cushion because, I could write you such a properly crazy and unnecessary story now about how like **because** [-] otherwise I couldn't afford it and how I have such a sad life blablabla oopsy […]')

There are some useful guidelines suggested in both previous research and Section 6.1 to determine whether a subordinating conjunction that introduces a paratactic clause functions as a coordinating conjunction or discourse marker. However, distinguishing between coordinating conjunctions and discourse markers in such instances is ultimately reliant on contextual clues and, based on the data of the *NottDeuYTSch* corpus, is often open to interpretation.

### 6.2.3 Modelling the grammaticalization of subordinating conjunctions

Having established that subordinating conjunctions have undergone grammaticalization where they may function as both coordinating and discourse markers in digital youth language, this section analyses how grammaticalization has previously been modelled. The section firstly examines the work of Gohl & Günthner (1999) and Günthner (1999), who developed a model based on the analysis of *weil* and *obwohl* in their data. Using the data from the *NottDeuYTSch* corpus and the guidelines identified in Sections 6.2.1 and 6.2.2, I then present a refinement to the model of grammaticalization that describes the emergent grammatical processes in digital youth language.

In the model developed by Gohl & Günthner (1999) and Günthner (1999), it is argued that the subordinating conjunctions *weil* and *obwohl* have undergone two stages of grammaticalization: first they were grammaticalized as coordinating conjunctions, then underwent subsequent grammaticalization as dis-

course markers. This suggests a linear model of grammaticalization, which I have adapted into a visual model in Figure 6.1 with examples.

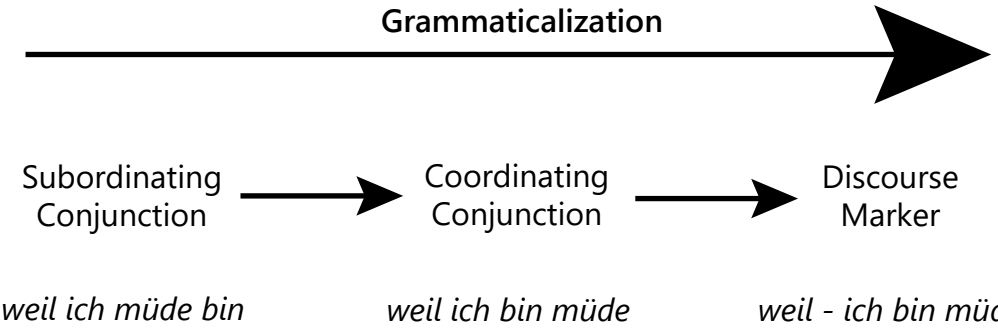



Figure 6.1: Linear model of grammaticalization of subordinating conjunctions, adapted from Gohl and Günthner (1999: 70)

However, there are problems with applying the model suggested by Gohl & Günthner (1999) to the data in the *NottDeuYTSch* corpus. The corpus contains examples of discourse markers that are identical to subordinating conjunctions, but which do not occur as coordinating conjunctions, such as *als ob* and *trotzdem* in Examples 6.28 and 6.32. This suggests that, contrary to the linear process in Gohl & Günthner (1999), a lexical item does not necessarily have to first be grammaticalized as a coordinating conjunction before functioning as a discourse marker. Rather, as shown in Figure 6.2, the data suggest that a subordinating conjunction can undergo grammaticalization to become *either* a coordinating conjunction *or* a discourse marker. Subsequently, a subordinating conjunction may undergo a second stage of grammaticalization, e.g. first being used as a coordinating conjunction and then as a discourse marker (which is the case with *weil*). On the data available, it is not possible to confirm that all lexical items that undergo the two-stage process this way. Indeed, some subordinating conjunctions may first undergo grammaticalization enabling them to function as a discourse and then later undergo an additional process to function as a coordinating conjunction.

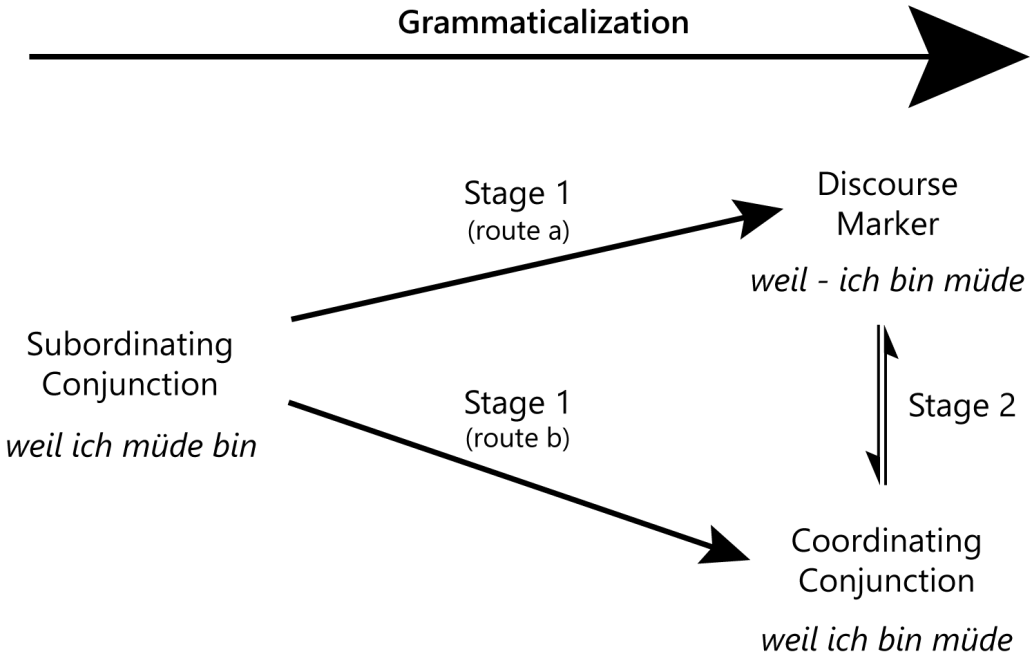



Figure 6.2: Refined model of grammaticalization of subordinating conjunctions

The usage of the 'postpositioner'[3] *wobei* in the corpus provides further evidence for the updated model presented in Figure 6.2. In the *NottDeuYTSch* corpus, there are 23 instances of *wobei* used as a discourse marker out of 800 total occurrences, as in Examples 6.25. However, there are no occurrences of *wobei* introducing parataxis in a subordinating clause.

**Example 6.25** (2014)
**wobei** MFSU [object from the computer game, MineCraft] is das schwierigste aber das kriegt ihr auch noch geschissen :D
('**actually** MFSU is the most difficult but you'll get it shite :D')

In light of this evidence, I contend that *wobei* functions as a discourse marker in the instances where it occurs in examples in several studies on main clause structure following subordinating conjunctions (see Günthner 2000b, 2002; Breindl, Volodina, & Waßner 2014; Freywald 2016). The *wobei* instances in Examples 6.26 and 6.27, taken from Günthner (2000b: 327) and Freywald (2016: 334), function as a conversation structuring signal, or more precisely, a "Reformulierungs-

[3] Defined in Chapter 5, Section 5.1.1.

marker" ('reformulation marker') (Blühdorn, Foolen, & Loureda 2017: 24; following Martín Zorraquino & Portolés 1999), which allows the speaker to correct or specify. In Example 6.26, the use of *wobei* is also syntactically non-standard as it introduces a sentence, i.e. "anteponiert" (see Chapter 5, Section 5.1.1), violating the expectation that a *wobei* clause must follow the main clause.

**Example 6.26** (adapted from Günthner (2000: 327))
23 Eva: des is nämlich NOCH schlimmer wie der wohnt,
24 weil der nämlich noch AUßERhalb wohnt.
25 hier ( ) bist ja wenigstens im paradies.
26 (1.5) ((Essengeräusche))
27 Karl: ja.
28 (1.5) ((Essengeräusche))
29 Karl: **wobei** AUßERHALB (-)
30 (das=is) in konstanz natürlich.
31 also (-) was heißt das schon (.) außerhalb.

('23 Eva: that is namely EVEN worse how he lives,
24 because namely he still lives OUTside [of a place].
25 here ( ) at least you're in paradise.
26 (1.5) ((eating sounds))
27 Karl: yes.
28 (1.5) ((eating sounds))
29 Karl: **well** OUTSIDE (-)
30 (that=is) in konstanz of course.
31 so (-) what does that actually mean (.) outside.')

**Example 6.27** (Freywald (2016: 334))
wenn die haare im gesicht hängen ..
das nervt wie sau ..
gleich nehm ich n haargummi .. **wobei**. nö das sieht scheiße aus

('if your hair hangs down in your face ..
that's annoying as hell ..

I'm going to get a hairband in a moment .. **actually**. nah that looks crap')

Only *dass*, *obwohl*, *weil*, and *wenn* (and *wenn* clusters) occur both as discourse markers and as coordinating conjunctions in the *NottDeuYTSch* corpus, while *als ob*, *soweit*, and *trotzdem* occur as discourse markers but not as coordinating conjunctions. All of these conjunctions are illustrated in Examples 6.28 to 6.35 below (discourse markers appear in bold and are followed by a dash (-) in the English translation).

**Example 6.28** (*als ob*: 2016)
**als ob** jetzt kann man dich ernst nehmmen also ich zu mindest
('**As if**[-] now people can take you seriously well me at least')

**Example 6.29** (*obwohl*: 2017)
[…] Und ja ich habe schon den ein oder anderen PC zusammengebaut… Ich hoffe mal nach deinem Kommentar. dass du es nicht gemacht hast… **obwohl** naja ist ja schließlich dann nur dein Geld.
('And yes, I have already assembled the odd PC or two… I hope after your comment that you haven't done it… **although** [-] well after all it's only your money.')

**Example 6.30** (*selbst wenn*: 2015)
[…] Was DayZ [computer game] angeht da hoffe ich jetzt auch das durch das neue Update es besser wird mit Hackern aber **selbst wenn** es gibt zumindest jetzt die Admin Konsole wo man nachverfolgen kann wer wenn wo getötet hat […]
('[…] When it comes to DayZ there I hope now as well that due to the new update it will be better with [the issue of] hackers but **even if not** [-] at least there's the admin console where you can keep track of who killed who where […]')

**Example 6.31** (*soweit*: 2014)
[…] Okay **soweit** ich bin ja nicht das einzige Kind, bei dem sich die Eltern trennen […]

('[…] Ok **so far** [-] I am of course not the only child whose parents have split up […]')

**Example 6.32** (*trotzdem*: 2012)
[…] sorry, aber **trotzdem** ich finds mutig ,dass ihr euch vor die Kamera traut um ,dass zu machen :'D !
('Sorry, but **anyway** [-] I find it brave ,that you have the guts to be on camera to, do that :'D !')

**Example 6.33** (*und wenn*: 2017)
+[COMMENTER] **und wenn** er ist halt so nirmand kann jemanden ZWINGEM SICH ZU ÄNDERN DU VOLLPFOSTEN😐
('+[COMMENTER] **and even if** [-] he is just like that no-one can force SOMEONE ELSE TO CHANGE YOU DIMWIT😐')

**Example 6.34** (*weil*: 2017)
Ich will das [YOUTUBER]-kissen **weil**, ich könnte dir jetzt so eine richtig bescheuerte und unnötige Story schreiben so von wegen ja sonst könnte ich es mir nicht leisten und ja ich hab so ein trauriges leben tralala hop-sasa […]
('I want the [YOUTUBER] cushion because, I could write you such a properly crazy and unnecessary story now about how like **because** [-] otherwise I couldn't afford it and how I have such a sad life blablabla oopsy […]')

**Example 6.35** (*wenn auch*: 2017)
beruig dich mal wieder, hab noch nie gesehen, dass jemand diese frage gestellt hat. und **wenn auch** dann hab ich ihn auf jeden fall nicht von dieser person geklaut, denn der ist mir nachdem ich alien [1979 film] geschaut hab spontan eingefallen.^^——>zuerst informieren, bevor man behauptungen ausspricht!
('calm yourself down, never seen somone who asked this question. and **even if** [-] then) I definitely never pinched it from this person, for it came to me spontaneously after I watched Alien.^^——>first inform yourself

before making assumptions!')

To sum up, the difference between a coordinating conjunction and a discourse marker is, then, not always clear, in part due to the frequent non-standard linguistic features in the *NottDeuYTSch* corpus. However, there is evidence in the corpus of conjunctions undergoing grammaticalization that have not been previously identified in previous research. We have seen that *dass*, *obwohl*, *weil*, and *wenn* can function both as a coordinating conjunction and as a discourse marker in addition to their standard function as subordinating conjunctions. Grammaticalization also occurs differently for different conjunctions. The evidence is compatible with a two-stage grammaticalization process for *weil*, functioning as both a coordinating conjunction and as a discourse marker. By contrast, for other conjunctions, such as *als ob*, there is evidence only of grammaticalization where it functions as a discourse marker, without an intermediary stage as coordinating conjunction, and I have proposed refining the model of Gohl & Günthner (1999) and Günthner (1999), accordingly.

## 6.3 Current and future trends for subordinating conjunctions

The analyses of *NottDeuYTSch* corpus data presented here demonstrate significant developments in the grammaticalization of subordinating conjunctions which allows them to function as both coordinating conjunctions and discourse markers. Furthermore, the chapter has presented methodological strategies to identify the function of the grammaticalized elements.

In Section 6.1, four contending alternative grammatical processes – topicalisation, novel conjunction clustering, multiple fronting, and the grammaticalization of the subordinating conjunction *weil* as a discourse marker – were examined to test how well they account for the relatively high frequencies of predicative elements, e.g. *dann* and *eigentlich*, in the first field position of clauses with non-standard non-subordinating word order. While all four processes may be valid, the process most likely to account for the data is the grammaticalization

of *weil* introducing the clause where it functions as a discourse marker, as *weil* in these cases functions to structure the information in the comment, rather than to imply causality between the clauses. As I have argued, the function of *weil* as a discourse marker is particularly apparent in cases followed by a prepositional phrase introduced by *wegen*. *Wegen* here already contains the causal link between the noun phrase and the rest of the clause, making the causal function of *weil* as a subordinating conjunction redundant, so instead *weil* introduces additional information in these cases, one of the functions of a discourse marker.

Further evidence of the grammaticalization of other subordinating conjunctions which allows them to function as coordinating conjunctions and discourse markers was presented in Sections 6.2.1 and 6.2.2 using new data from the *NottDeuYTSch* corpus. Here I tested a range of hypotheses drawn from research to date on the potential reasons for the occurrence of subordinating conjunctions introducing paratactic clauses. The data in the *NottDeuYTSch* corpus do not provide clear support for any of the previous hypotheses (explored in Section 6.2.1). Rather, the use of the grammaticalized forms of subordinating conjunctions seems to have generalised as part of the wider adoption of the phenomenon within digital youth language. Previous explanations for the occurrence of this phenomenon are no longer necessarily applicable to German youth language, although they might have been valid at the time of writing, for example, the hypothesis that paratactic *weil* clauses occur when *weil* is analogous to *denn* (see Reis 2013).

There is also evidence in the *NottDeuYTSch* corpus of the grammaticalization of six additional subordinating conjunctions (*als ob*, *selbst wenn*, *soweit*, *trotzdem*, *und wenn*, and *wenn auch*) which allows them to function as discourse markers without evidence of first functioning as coordinating conjunctions. In previous studies (e.g. Freywald 2016), only *wobei* has been identified in such a role. Furthermore, the *NottDeuYTSch* corpus data provides evidence of the function of *dass* as a coordinating conjunction, which has previously only been identified as functioning as a discourse marker when introducing paratactic clauses (Freywald 2016).

In light of the emergent grammatical developments, a more complex grammaticalization model is necessary than previously proposed, as I argued in Section 6.2.3. The linear model of Gohl & Günthner (1999) and Günthner (1999) – according to which a subordinating conjunction first undergoes grammaticalization where it functions as a coordinating conjunction, before a second stage where it functions as a discourse marker – cannot account for all the data in the corpus. Rather, there is evidence to suggest that subordinating conjunctions may undergo grammaticalization where they function directly as discourse markers without first functioning as coordinating conjunctions. However, there is a chance that this is due to a gap in the data, so this analysis should be test on an even larger dataset to verify the tentative findings in this study. It remains to be seen whether, in future, the subordinating conjunctions that function in that subordinating role and as discourse markers may be subsequently used as coordinating conjunctions. Nevertheless, it is clear that the paratactic use of subordinating conjunctions, regardless of function, has become increasingly accepted by young people in digital youth language over the short ten-year timeframe covered by the *NottDeuYTSch* corpus, and that similar syntactic change has spread to other subordinating conjunctions.

# Part III

# Orthographic innovation in Digital Writing

# 7

# *Deine rechtschreibung ist god like kappa :)* – the use of orthographic features for metacommunication

Chapters 4 to 6 examined lexical and syntactic developments in digital youth language. This chapter turns to how meaning-making in Digitally Mediated Communication (DMC) can occur at the orthographic level to alter the meta-communicative information of a message, i.e. how a message may be received und understood by the reader through the use of orthographic innovations to convey emotion, tone, as well as other information about the author of the digital text. These orthographic features can include the selection of individual characters, including graphemes, spaces, and other non-alphanumeric characters, such as punctuation and emoji. For example, writing a word entirely in capital letters, as shown in Example 7.1, is widely accepted in digital writing as a visual equivalent of increased volume in speech (popularly referred to as 'shouting capitals,' Crystal 2006: 37). However, the metacommunicative functions of orthographic features are not restricted to the imitation of spoken lan-

guage. For example, using capital letters can also intensify the emotion conveyed (Parkins 2012: 48; Hilte, Vandekerckhove, & Daelemans 2019: 297), as in Example 7.1, where the commenter perhaps wishes to emphasise the emotion they feel when talking about their "biggest dream". Alternatively, shouting capitals can be used to highlight the semantic salience of part of the message (McAteer 1992: 350), in this case the possibility of meeting the *YouTuber*.

**Example 7.1** (2016)
mein GRÖßTER Traum: DICH TREFFEN😍💘😍[...]😘😘😍 ('my BIGGEST dream: MEETING YOU😍💘😍[...]😘😘😍')

The Chapter is split into two main sections. Sections 7.1 and 7.2 provides a methodological background for the analysis of metacommunication and detail the approaches taken to identify the data for subsequent analysis. Section 7.1 collates and contextualises orthographic features of digital writing identified in research to date, as well as innovative features in the *NottDeuYTSch* corpus, highlighting the socio-technological developments within DMC over the past 40 years. Section 7.2 then describes the computational linguistic methods used to capture the groups of orthographic features used to carry metacommunicative functions in the *NottDeuYTSch* corpus. These groups of orthographic features are then summarised in Section 7.3. The resulting data inform the holistic framework of metacommunicative functions presented in Section 7.4, to yield a single, robust system of analysis that can be applied to all non-standard orthographic features in Digital Writing.

## 7.1 Orthographic features in DMC

I have collated the many different orthographic features identified in scholarship from 1990-2018 (as discussed in Chapter 2, Section 2.1.1) into three broad categories, related to how they are either displayed within digital spaces or alter existing linguistic elements. This ensures both that the categories will remain valid even if the features change over time, and that the framework for the analysis of metacommunication presented in Section 7.4 remains applicable

to future technological developments in text-based DMC as much as possible. The three categories are:

1. Graphicons (e.g. emoji < 😊 >)
2. Creative orthography (e.g. shouting capitals < HALLO >)
3. Metalinguistic signs (e.g. hashtags < #GlockeAktiv >)

Below, I describe each of these three categories in turn.

### 7.1.1 Graphicons

The first category of orthographic features, graphicons, includes emoticons, emoji, kaomoji, and ASCII art. These features are used in similar ways to add pictorial information to a message, but differ in visual appearance and in how they are constructed. Descriptions, together with examples of how each type of feature can portray a smiling face, are listed in Table 7.1.

The term 'graphicon' was suggested by Herring & Dainas (2017: 2185) to include "emoticons, emoji, stickers, GIFS, images, and videos". In this chapter, I use the term to refer to all combinations of characters used to create pictographic or ideogrammatic symbols that can be integrated with text-based language in a single digital message, excluding visual elements, such as GIFs or pictures that are sent in a separate space to the text. Other competing terms for this category of features include "Bildzeichen" (Dürscheid & Siever 2017; Androutsopoulos 2018; Siebenhaar 2018; Beißwenger & Pappert 2019a, 2020), "graphic characters" (Miyake 2007), "ideograms" (Giannoulis & Wilde 2019; Lu et al. 2016), "pictograms" (Hinz 2015; Danesi 2019; Giannoulis & Wilde 2019), and "pictographs" (Pavalanathan & Eisenstein 2015; Sampietro 2016; Hougaard & Rathje 2018). Stark & Crawford (2015: 5), who consider that emoji, emoticons and kaomoji "straddle the line between ideogram and pictogram", define ideograms as "symbolic representations of a particular concept or idea", and pictograms as "ideograms that show a pictorial image of the object being represented". However, graphical elements used in DMC can contain both pictorial or ideogrammatic aspects, which can index a wide variety of meanings and functions depending on the context, as we shall see in the next chapter. I therefore

Table 7.1: Types of graphicon

| Graphicon Type | Description | Example |
|---|---|---|
| Emoticons | Typographic approximations of faces, body parts rotated 90 degrees. Primarily formed by combining punctuation and alphanumeric characters. Users are free to construct their own emoticons using these resources | :) |
| Emoji | Small graphical representations of faces, people, things, ideas, and concepts. Emoji are designed and defined by the Unicode consortium and are then rendered by various operating systems. Users can freely choose combinations of emoji but cannot design their own | |
| Kaomoji | Typographic approximations of faces, similar to emoticons, but not rotated. This style is popular in East Asia and kaomoji sometimes contain characters available in Hiragana, Katakana or other Eastern Asian syllabaries or alphabets | ^_^ |
| ASCII art | More complex graphic design originally created from ASCII characters (the alphanumeric characters, symbols, and punctuation available in early computing). The term now includes graphical art that uses any available Unicode symbol. Emoticons can be said to be the simplest form of ASCII art | |

prefer the term "graphicons", as it suggests the form of the features but makes no assumptions about their potential functions.

Early research on graphicons used in DMC mainly focused on emoticons, e.g. the disappointed face in Example 7.2, and kaomoji, such as < ^^ > to represent smiling eyes or eyebrows, as in Example 7.3 (e.g. Herring 1996a; Witmer & Katzman 1997). However, early studies rarely examined more complex kaomoji that incorporated Japanese characters, e.g. < (つ◕_◕)つ >, in which the hiragana character, つ [zɯ] represents hands giving a hug (the 'dakuten' diacritic ゛ can be interpreted as either fingers or indicating motion).

**Example 7.2** (2013)

Das ist das schlechteste video von euch :/ ('That is the worst video from you :/')

**Example 7.3** (2013)
echt cool ^^ ('really cool ^^')

ASCII art evolved concurrently with emoticons and kaomoji, but has received little attention in linguistic DMC scholarship. It features sporadically as an addendum to analyses of emoticons (e.g. Haase et al. 1997: 78; Bader 2002: 95) or digital writing practices (e.g. Paolillo 2011: 6), but is often omitted from studies on other graphicons (e.g. Table 1 in Siebenhaar 2018: 753). However, scholarship on ASCII art is present in more design-oriented fields (e.g. Thaler 2003). In fact, ASCII art can have several metacommunicative functions, as we shall see in Section 7.4.

Despite being available in mobile phone communication since 1997 (Burge 2019a), emoji were not easily available nor fully compatible between different operating systems until they were standardised as part of Unicode 6.0 in October 2010 (The Unicode Consortium 2011). From then on, emoji began to replace emoticons and kaomoji in online communication. In the *NottDeuYTSch* corpus, emoji usage overtook emoticons and kaomoji in 2015 (see Figure 7.1). Whilst the first research into emoji usage started in Japan in the mid-2000s (see Miyake 2007), it was not until the next decade that similar research began in Europe and North America. Researchers sought to create taxonomies of emoji functions (e.g. Ge & Gretzel 2018; Gawne & McCulloch 2019; Beißwenger & Pappert 2019a), or to examine particular sets of functions, e.g. the expression of feelings (Hougaard & Rathje 2018), conversational management (Sampietro 2019), word or grapheme replacement (Dürscheid & Siever 2017), or use as punctuation (Sampietro 2016; Siebenhaar 2018). Other work investigated: the negative correlation between the frequency of emoji use within digital writing and the frequency of emoticons (e.g. Pavalanathan & Eisenstein 2016); the place of emoji in the broader history of similar features, including emoticons and stickers (Konrad, Herring, & Choi 2020); and the demographics of emoji users (e.g. Dürscheid & Siever 2017; Siebenhaar 2018; Hilte, Vandekerckhove, & Daelemans 2019). However, there is little research on how emoji are used in combination with orthographic features other than graphicons (one exception

being Suttles & Ide 2013).

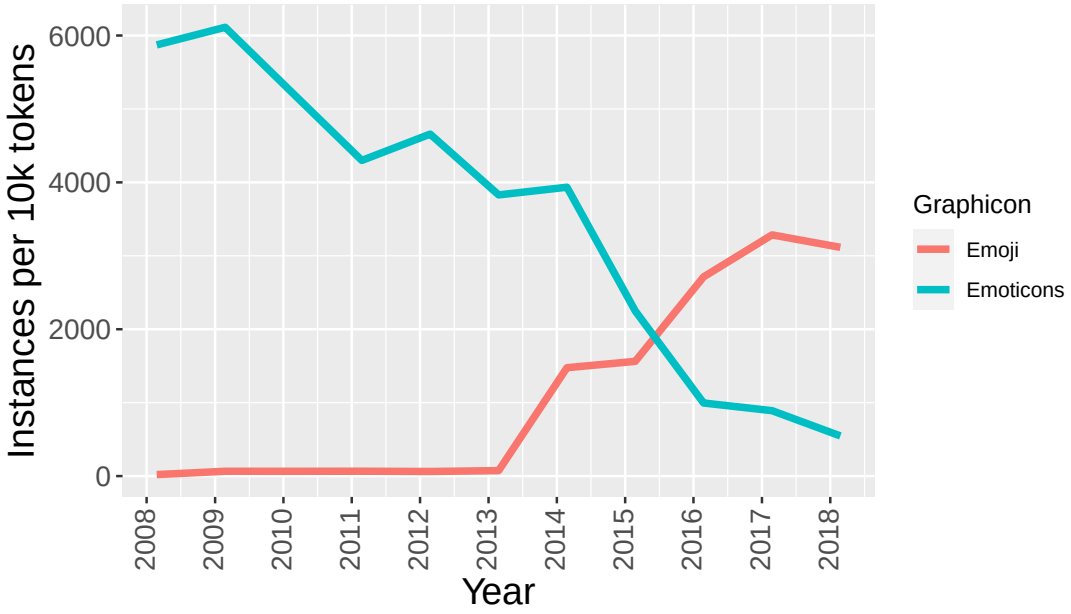



Figure 7.1: Usage of emoji and emoticons over the timespan of the *NottDeuYTSch* corpus

All four forms of graphicon, emoticons, emoji, kaomoji, and ASCII art, are present in the *NottDeuYTSch* corpus and are included in the framework of metacommunicative functions presented in Section 7.4.

### 7.1.2 Creative orthography

The category that I have labelled *creative orthography* includes the following groups of orthographic features:

1. The repetition of graphemes
2. The repetition of punctuation marks
3. The creative use of letter case
4. The creative use of spacing

**The repetition of graphemes and punctuation marks.** The repetition of punctuation and graphemes is found in DMC texts since the 1980s (e.g. Carey 1980; Werry 1996; Hentschel 1998), and the feature is used chiefly to imitate the prosody of spoken language or to otherwise draw attention to the word(s).

In the case of Example 7.4, the repetition of the < a > in < jaaa > can be interpreted as the imitation of a lengthened vowel sound in speech, which can express "strong social or emotional involvement" (De Decker & Vandekerckhove 2017: 255). The repetition of full stops is normally used to suggestion omission or incompletion, i.e. they allow "dialogic expansion" (Vandergriff 2013: 8), although they are used metacommunicatively to suggest irony (Thompson & Filik 2016: 116). Both of these interpretations can be applied to Example 7.4: either the commenter is hinting at something unspoken or uses the full stops to indicate that the < jaaa > is not seriously meant.

**Example 7.4** (2010)
Das wär geil jaaa….. ('That would be cool yeaaah…..')

**Creative use of letter case.** The use of non-standard letter case has diversified from the early days of DMC. Initially, writing words (or parts of words) in capital letters was one of the only metacommunicative orthographic features used for "stress" (Carey 1980: 68), as shown in Example 7.5, where the comment, written entirely in capital letters, can be interpreted as having been 'shouted'.

**Example 7.5** (2016)
ICH WILL WISSEN WO DAS HAUS STEHT !!!????? ('I WANT TO KNOW WHERE THE HOUSE IS !!!?????')

A notable development in the creative use of letter case in DMC is the alternation of upper and lower-case letters (e.g. 'kOmpllmenT,' Androutsopoulos 2007: 84), which Androutsopoulos (2003a: 186) refers to as the "writing style of hackers and crackers". This style was closely associated with so-called "Leet Speak" (from 'elite' speak), the replacement of letters "by nonalphabetic symbols based on graphic resemblance" (e.g. *leet* could be written as "1337"), originating amongst hacking communities (Herring 2012: 2).[1] The use of alternating capital letters and the replacement of certain letters with numbers is found in the *NottDeuYTSch* corpus, but mostly either in the usernames of commenters

[1] Leet Speak and alternating capital letters differ from so-called 'Camel Case' (also referred to as 'Pascal Case' or 'Initial Caps'), where phrases are written without spaces, with the first letter of each word capitalised, e.g. 'PascalCase'.

(not investigated in this study) or, alternatively, used as parody, as shown in Example 7.6. However, the meaning of this feature seems to have changed in both wider internet usage and in the *NottDeuYTSch* corpus since 2017, more recently indicating a mocking, ironic effect, as shown in Example 7.7 from 2018. The change in usage seems to have stemmed from the grapheme stylisation in a meme involving a picture of the cartoon character, Spongebob Squarepants, imitating a chicken (see Figure 7.2), which often accompanies such stylised text in mixed media spaces, such as Facebook or Reddit (Know Your Meme 2021).

**Example 7.6** (2011)
iCh haSze lEut3 diie So sChreib3n -.-' (i haTe pEoPl3 whoo wRiT3 LiK3 ThIs -.-')

**Example 7.7** (2018)
DiE KLaUn UnSRe jO[B]S!!2!1!1!!! (ThEy'Re sTeAlIn OuR jO[B]S!!2!1!1!!!)

**Creative use of spacings.** The creative "spatial arrangement" of words within the messaging space was noted by Carey (1980: 68), but the study of such features has not featured prominently in subsequent DMC research – only two features have been reported in previous work: 'line spacing', i.e. comments containing multiple line breaks; and 'letter spacing', the use of spacing between letters in a word. Both of these features can have a metacommunicative function, which we shall see in Section 7.4. Letter spacing has also been documented by Hentschel (1998: 3.2.4) and Crystal (2006: 92), who regard it as a form of emphasis; but, again, it has rarely been mentioned in scholarship. The only other brief mention of it within a DMC context briefly occurs in an analysis of sound symbolism in DMC (see Jurčević 2019: 98).

### 7.1.3 Metalinguistic signs

A third category of orthographic features consists of inflectives, hashtags, representations of laughter, and the literal representations of visual metacommunication. Inflectives (*Inflektive*) are verb stems enclosed in asterisks or other punctuation (see the use of < *grins* > in Example 7.8), that are "direct repre-

Figure 7.2: Mocking Spongebob Meme with accompanying stylised text

sentations of actions or inner states" (Schlobinski 2001: 204). However, the text of an inflective can contain more syntactically complex verb phrases, e.g. intensifiers, adverbs or nouns that precede the verb stem (see Table 7.2).

> **Example 7.8** (from Schlobinski (2001: 203))
> oma [Username] ist mnemo [Username] schon wieder da oder noch? *grins* ('oma is mnemo already back or still? *grin*')

Inflectives featured prominently in studies of German-language DMC between the 1990s and early 2010s Dürscheid, Wagner, & Brommer (2010), but the use of inflectives has fallen dramatically in the *NottDeuYTSch* corpus, as shown in Figure 7.3. In 2008, there were 4,903 inflectives per million tokens in the corpus, which, for reference, is a third more frequent than in the *Dortmunder Chat*

Table 7.2: Complex inflectives, adapted from Schlobinski (2001: 193)

| Inflective (original) | Inflective (English) | Syntactic construction |
|---|---|---|
| knuddel | cuddle | Verb stem |
| zurueckknuddel | cuddle back | Verb particle + verb stem |
| megaknuddel | mega cuddle | Intensifier + verb stem |
| megazurueckknuddel | mega cuddle back | Intensifier + verb particle + verb stem |
| warriorknuddel | warrior cuddle | Noun + verb stem |
| siskyauchknuddel | cuddle sisky too | Noun + adverb + verb stem |
| dich ganzdollknuddel | cuddle you really hard | Pronoun + adverb + adjective + verb stem |

*Korpus* of online conversations between 1998-2006 (3,407 instances per million tokens). However, inflective usage saw a steep decline in the *NottDeuYTSch* corpus between 2008 and 2011 (a fall from 4903 ipm to 573 ipm), before dropping almost completely out of use by 2016 (118 ipm).

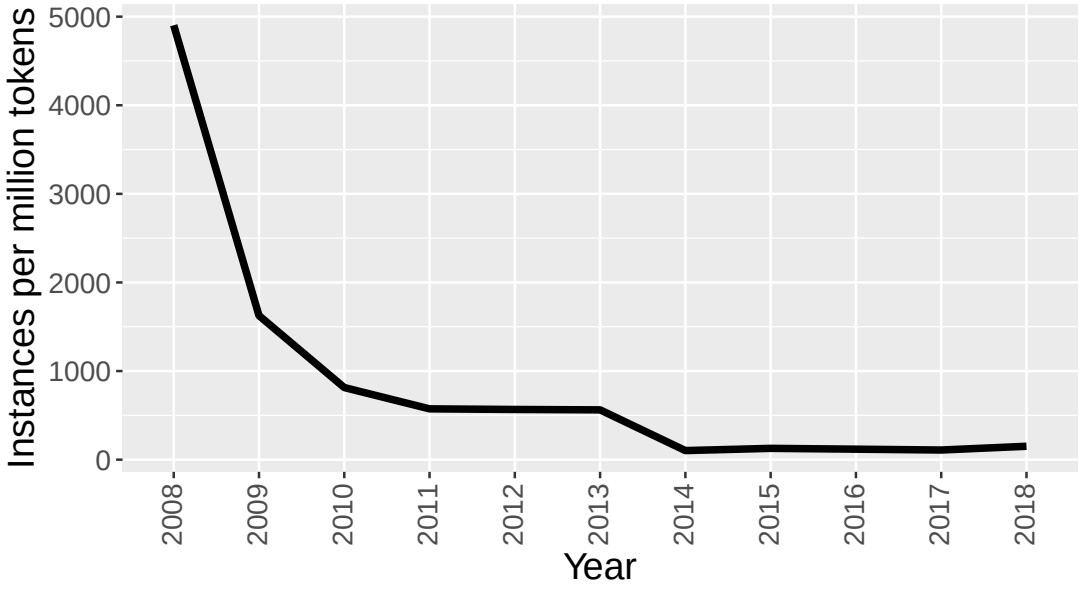



Figure 7.3: Frequency of inflectives over time in the *NottDeuYTSch* corpus

The second feature categorised within the category of metalinguistic signs is the use of hashtags. These are orthographically similar to inflectives, but the text is preceded by a hash symbol (< # >, also called an octothorpe), as shown in Examples 7.9 to 7.12, rather than enclosed by asterisks. Hashtags were popularised in

late 2007 on Twitter, initially as a way of tagging keywords for cross-referencing, and were later adopted by other social media sites (K Scott 2015: 9). In addition to the basic labelling function, hashtags were soon used as "metacommunicative tagging" (Daer, Hoffman, & Goodman 2015: 13) to signify the emotion of the author or to provide other meta-information about the message, such as a "taxonomic classification" or a "meta-evaluation" of the message (Zappavigna 2015: 278). Unlike inflectives, the text of a hashtag is not restricted to verb stem-based constructions, and can include a wide variety of syntactic constructions, from nouns (Example 7.9), to phrases (Example 7.10), to entire sentences (Example 7.11). In multi-word hashtags, the spaces between words are removed, as otherwise only the first word would be 'hashtagged', i.e. function as a keyword. In the Twitter data used by Zappavigna (2015), hashtags were integrated within a sentence, so that they had both a metacommunicative and literal function, as shown in Example 7.12. This usage is rare in the *NottDeuYTSch* corpus (Example 7.11), with hashtags primarily unintegrated, either at the end of clauses or sentences (Example 7.10), or functioning similar to interjections (Example 7.9).

**Example 7.9** (2017)
#FAIL KOMM DOCH NACH BAYERN ☺ ich kann dir meins geben haha ('#FAIL JUST COME TO BAVARIA ☺ I can give you mine haha')

**Example 7.10** (2018)
[YOUTUBER] ABI BENI GRÜSSEN YAP NÄCHSE VIDEO DA #GLOCKEAKTIV ('[YOUTUBER] BROTHER GREET ME ALSO MAKE NEXT VIDEO HERE #BELLACTIVE [Note: This comment uses both Turkish and German]')

**Example 7.11** (2017)
Ich würde das Teil verdammt gerne Haben #ichwilldasding für ps4 ('I would damn well love to have that part #iwantthething for ps4')

**Example 7.12** (adapted from Zappavigna (2015: 277))
Why do we love #Obamacare? Here's 47 million great reasons

In Example 7.10, the hashtag *#GLOCKEAKTIV* can be interpreted as signalling

the commenter's approval, expressed by pressing the notification bell to signal their support of the channel. In Example 7.11, despite the hashtag comprising a grammatically complete sentence, the primary function is metacommunicative, i.e. to tag the comment (the original function of a hashtag, as noted by K Scott 2015) to be identified by the *YouTuber* as a competition entry, where a comment picked from those that used the specific hashtag would win a prize.

In addition to conventionalised characters, such as the hash symbol and asterisk, which designate specific text as metacommunicative, there is evidence that words, acronyms, and initialisms have also been conventionalised to be interpreted as primarily metacommunicative rather than literally. One such category of metalinguistic signs is used for the representation of laughter, both through onomatopoeic representations, such as < haha >, and acronyms or initialisms, such as < lol > (laughing out loud) or < rofl > (rolling on the floor laughing). In addition to < haha >, < lol >, and < rofl >, I have found several variants both in the *NottDeuYTSch* corpus and identified in other research on English and German online language (e.g. Crystal 2004; Diekmannshenke 2007; Barton & Lee 2013; Bahlo et al. 2019). These variants include < lulul >, < roflcopter >, and < hihihi > (see Example 7.13). Much of the previous research has established that these features have primarily phatic functions (Baron 2004; Tagliamonte & Denis 2008) and turn-organising or repairing functions (Petitjean & Morel 2017; Tagg 2009), rather than being a literal representation of the sounds made by the commenter, a fact which prompted Crystal (2006: 37) to ask "how many people are actually 'laughing out loud' when they send *LOL*?".

**Example 7.13** (2015)
Eure arme blume hihihi du bist die beste [YOUTUBER]😂😂😂 ('Your poor flower hihihi you are the best [YOUTUBER]😂😂😂')

The fourth group of metalinguistic signs is the literal representation of visual metacommunication, such as 'spelling out' emoticons, e.g. writing < icks de >, instead of < xD >, as in Example 7.14 (also see Langenscheidt 2016). This kind of interplay between spoken and written language is an important source of metacommunicative creativity. Many of the internet-based neologisms may

have started as initialisms (i.e. each letter was referred to individually), but were soon pronounced as a single word (an acronym), e.g. [lɒl] for < lol >, rather than [ɛl.əʊ.ɛl], when referred to in spoken communication (Crystal 2011: 61). This incorporation of initialisms into spoken language established a feedback loop that has in turn created other metalinguistic signs. For example, the letters of < lol > are phonotactically plausible in English and German, but other less plausible initialisms were also adapted into acronyms, often for humorous effect, such as < lmao > [ləmaʊ] (laughing my ass off) or < rofl > [rɒfəl] (Ulaby 2006). Another metalinguistic sign involving respelling is the replacement of exclamation marks with < eins > or < elf >, which can refer to a commenter accidentally typing < 1 > instead of pressing the "shift" button and "1" to write < ! >, as shown in Example 7.15, where the commenter develops the joke further by writing < zwölf > ('twelve'). Blashki & Nichol (2005) state that this feature is part of the humour of 2000s "game geek" subculture, as part of leet speak.

**Example 7.14** (2010)
jo jo voll geil alta icks de ('yo yo so cool man ecks dee')

**Example 7.15** (2014)
MEINER!!!!!!!!!!!1111!!!1!elf!1!zwölf! ('MINE!!!!!!!!!!!1111!!!1!eleven!1!twelve!')

Much research has been devoted to identifying emotion and tone in a message, particularly sarcasm and irony (e.g. Davidov, Tsur, & Rappoport 2010; Thompson & Filik 2016), when "literal and intended meanings are in opposition" in sarcastic messages, since identifying the intended tone of written messages can pose a challenge for sentiment analysis (Barbieri, Saggion, & Ronzano 2014: 50). Conventions that can be used to convey irony, such as emoticons or emoji that either wink or display a tongue sticking out, e.g. < ;) >, < :P >, < 😉 >, and < 😜 > (see Thompson et al. 2016; Weissman & Tanner 2018), are not always reliable indicators of irony, as the symbols are open to other interpretations. In addition to emoticons and emoji, on the video game streaming website, *Twitch.tv*, commenters can use 'emotes', small in-line images used to convey a wide spectrum of metacommunicative functions: from more widely-used concepts, such as smiley faces similar to emoji, to emotes that are specific

to *Twitch* channels (and which sometimes require a paid subscription to use), such as the 'gronkhHi' emote in Figure 7.4, a cartoon representation of the German *YouTuber* and streamer, Gronkh, the use of which is exclusive to subscribers of his *Twitch* channel.

Figure 7.4: The 'gronkhHi' *Twitch* emote

Emotes on *Twitch* are produced either by selecting them from a menu or by typing the name of the emote enclosed by colons, e.g. < :gronkhHi: >. The text is then replaced in-line by the desired emote. Due to *Twitch*'s popularity amongst gaming communities, some gaming *YouTubers* publish their *Twitch* streams on *YouTube* as well, which fuels the cross-pollination of audiences between the two sites. This has led to commenters on *YouTube* referring to *Twitch*'s most famous emote, 'kappa' (see Figure 7.5), used to indicate that the message is intended to be interpreted ironically (see the entry on *kappa* in Langenscheidt 2016).

Figure 7.5: The 'kappa' *Twitch* emote and variants: kappaPride and kappaClaus

However, as emotes do not render as images on *YouTube*, commenters simply write < kappa > in their messages to achieve the same effect. In Example 7.16, in writing < kappa > at the end of the message, the commenter is indicating that they do not, in fact, think that the spelling of their conversational partner is 'god-like'. In the *NottDeuYTSch* corpus, the usage of < kappa > is primarily localised under videos in the 'Gaming' category, which reflects the shared audience with *Twitch*, but its usage has increased considerably within this category, which justifies its inclusion in the analysis, as it appears in 241 comments.

While all variants of < kappa > were included in the regular expression, such as < keepo >, < kappaPride >, and < kappaClaus >, only < kappa > appeared in the comments.

**Example 7.16** (2016)
Deine rechtschreibung ist god like kappa :) ('Your spelling is god-like kappa :)')

## 7.2 Capturing orthographic features in the *NottDeuYTSch* corpus

Having introduced the orthographic features identified in previous work that can be used to metacommunicatively alter the message in a digital text, this section outlines the methodological processes to identify and capture the selected groups of orthographic features in the *NottDeuYTSch* corpus in preparation for analysis in Section 7.4.

### 7.2.1 Identifying graphicons in the *NottDeuYTSch* corpus

Graphicons were identified in the *NottDeuYTSch* corpus using two different regular expressions. Capturing emoji was the most straightforward: using the latest list of emoji from the Unicode organisation (The Unicode Consortium 2021b), I extracted the code points (e.g. U+1F600 for the Grinning Face emoji, 😀) and arranged them so that emoji with Zero Width Joiners (used to combine multiple emoji to create new variants) would be captured first to prevent emoji being counted twice. An example of the combination of multiple emoji is "Woman Running: Light Skin Tone" (🏃🏻‍♀), which is comprised of four code points: U+1F3C3 (Person Running, 🏃), U+1F3FB (Light Skin Tone, 🏻), U+200D (Zero Width Joiner), and U+2640 (Female Sign, ♀).

Capturing emoticons, kaomoji, and ASCII art required significantly more work, as they are not single code points but combinations of common characters. I developed a five-stage iterative process to identify the common patterns used

in emoticons and kaomoji, based on sampling progressively larger random samples from the corpus. The five stages, labelled according to the kind of graphicons that they capture, are as follows:

**1. Heart emoticons:** A regular expression that captures the frequent emoticon, < <3 > to represent a heart (as well as < </3 >, which represents heartbreak).

**Regular Expression:** `(\\<[\\/\\]?3+)`

**2. Eyes-first emoticons:** A regular expression that captures emoticons that are rotated 90 degrees anti-clockwise so the characters used to represent the eyes precede the mouth, e.g. < :-) >.

**Regular Expression:** `([\:\;\|\^\=xXB8][\-\^\'\"]*([Vv3DdOoPp\|\^\@\$\*\\\)\(\/\[\]])\4+(?=\s|[[:punct:]]|$))`

**3. Mouth-first emoticons:** A regular expression that captures emoticons that are rotated 90 degrees clockwise so the characters used to represent the mouth precede the eyes, e.g. < D: >.

**Regular Expression:** `((?<!\w)([DdOoPp\|\^ \@\$\*\\\)\(\/\[\]]+)([\_\.\-\^\'\"]*)([\:\;\| \^\=xXB8\-])(?=\s|[\"\!\?\'\,]|$))`

**4. Kaomoji:** A regular expression that captures the combinations of characters where the emulation of faces are not rotated, e.g. < ^.^ >, including kaomoji that contain non-Latin characters, e.g. < (✿◠‿◠)つ >.

**Regular Expression:** `((?<=[A-z]|\s)(?:(?:[oO*^]+ [\.\-\_]+[oO*^]+ ))|(?:\^{2,})|(?:[-]+[\.\_]+ [-]+))`

**5. ASCII art and other constructions:** A regular expression that captures ASCII art and other complex emoticons that were not captured by the above four regular expressions, such as < (,,,)(>.<)(,,,) >.

**Regular Expression:** `((\w?([\=\\\/\[\]\()]+)(([YL8y_.]|[^A-z0-9?]+|[^A-z0-9?]+[\S]?[^A-z0-9?]+)([\=\\\/\[\]\)]+))\3*\w?)(?=\s|[xXBD])|(((?<=[A-z]|\s)(?:(?:[oO*^]+[\.\-\_]+[oO*^]+))|(?:\^{2,})|(?:[-]+[\.\_]+[-]+)))`

In some online German-speaking communities, such as some German-speaking Reddit forums, the character < Ü > is used to represent a grinning face and eyes. While this can be considered an emoticon, it is not included in this analysis because, out of the 112 total instances of < Ü > on its own in the corpus, it is never used as an emoticon, instead fulfilling other functions. For example, as shown in Example 7.17, it can appear as a direct quote of the letter, as an abbreviation of *über*, or as part of a word that has been stylistically separated through spacing.

**Example 7.17** (Usages of 'Ü')

2017: Du hast ausfersehen U ALS Ü geschrieben (You have accidentally written U as Ü)

2017: Ü 40?😂 (Over 40s?😂)

2012: Das Video is M Ü L L (The video is R U B B I S H)

### 7.2.2 Identifying creative orthography in the *NottDeuYTSch* corpus

The methodological procedures to identify features categorised as creative orthography were as follows:

For the repetition of graphemes and punctuation, three or more identical consecutive characters were counted as deliberate repetition, following the methodology of Hilte, Vandekerckhove, & Daelemans (2019: 296–97). Capturing just two repeated characters presents a significant methodological challenge, as it would require the removal of all words written with two consecutive identical graphemes (e.g. < cool >). Even if this were this to be easily achievable, the most important aspect of the feature is deliberate repetition in

order to provide a metacommunicative effect; two identical characters could simply be an orthographic oversight, rather than a deliberate strategy.

The use of creative letter case was captured by locating words written either all in capital letters, with three or more consecutive capital letters within the word, or which used alternative letter case in groups of up to two letters (e.g. < leT-Ter CaSe >). Acronyms and initialisms were then manually removed (e.g. *AFD* (1524 instances) and *HDGDL* (1127 instances)). However, the omission of capitalisation of the initial letter according to the rules of standard German has been excluded. While it can be a stylistic choice (Crystal 2006: 90; Tagliamonte & Denis 2008: 26), and therefore could have a metacommunicative function, it cannot be reliably used to interpret metacommunicative functions due to its primary usage as a time-saving measure, as part of what is referred to as "Sprachökonomie" ('language economisation', see Chapter 2, Section 2.1.1.1).

The features in the creative spacing category were captured using regular expressions according to the following conditions:

1. **Letter spacing:** Words of over three letters where the letters are separated by spaces.
2. **Line spacing:** Comments with two or more manually inserted line breaks.

Capturing comments with line breaks meant that some ASCII art was identified by the regular expression. These were not removed, as ASCII art often contains creatively spaced text either accompanying it or incorporated within it (see Figure 7.6), and such instances are analysed in this chapter both as a graphicon and as an example of creative spacing, due to their complex nature. The creative spacing regular expression also captured comments with features, such as lists and paragraphing, that could be considered a part of 'standard language' in other textual media. However, within informal digital writing settings, such as *YouTube* comments, text input is not normally paragraphed, partially as a result the fact that of messaging clients and social networks, such as *WhatsApp* (WhatsApp 2020) or *YouTube* (see Newgrounds 2012) send the message when pressing the 'Enter' key, rather than inserting a line break. I argue in Section 7.4.4, on the basis of evidence from the corpus, that the deliberate insertion of

new lines, achieved through keyboard combinations such as 'Control + Enter', is a departure from established DMC norms, and is an instance of metacommunication, in line with the research by Carey (1980) discussed in Section 7.1.2 above.

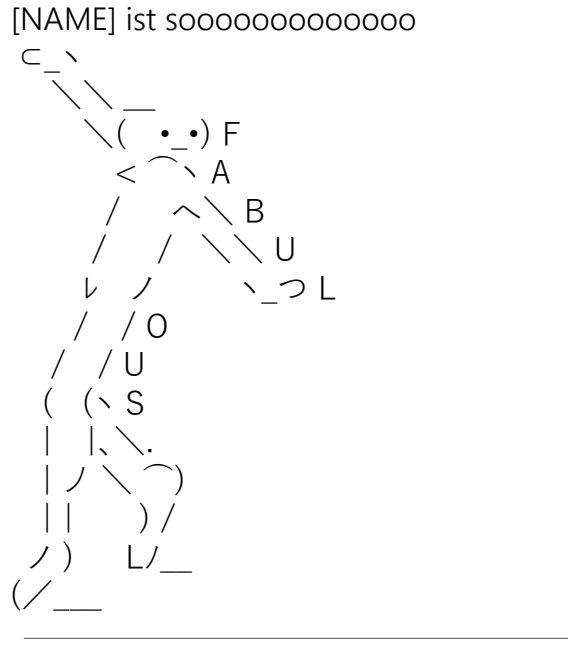



Figure 7.6: ASCII art of a person posing or dancing, 2016

### 7.2.3 Identifying metalinguistic signs in the *NottDeuYTSch* corpus

Features in the third category of metacommunicative orthographic features, metalinguistic signs, required the most methodological processing. Inflectives were captured using RegEx by searching for words and phrases delimited by < * >. In previous research, < + > has also been identified as a strategy to indicate an inflective, but this was not present in the *NottDeuYTSch* corpus. Other delimiters, including < ~ > and < / >, were explored, but these tended to be used to emphasise the words in the message (referred to as 'sparkle punctuation' in McCulloch 2019: 127), as part of a creative orthographic strategy to manipulate the space surrounding the word, which is demonstrated in Example 7.18. The manipulation of words within the comment space is examined in

further detail in Section 7.4.4.

**Example 7.18** (2018)
~LOL~

Hashtags were perhaps the most straightforward of the metalinguistic signs to capture, although there were some individual cases where the commenter had separated the words presumably intended to be prefixed by the hashtag with spaces (e.g. < #no click bait >), which meant the rest of the intended hashtag was not captured, only < #no >. One accepted alternative, present in the *NottDeuYTSch* corpus, is to replace the potential spaces by writing the word in so-called 'Camel case' to mark word breaks, as in Example 7.19 with the hashtagged phrase < #NetzFragtMerkel > (rather than < #Netz fragt Merkel >). This technique ensures that the hashtag links to the entire phrase.

**Example 7.19** (2015)
**#NetzFragtMerkel** was wird jetzt im Freihandels Abkommen zugelassen oder nicht und wen sie einen T-rex hätten wie würden sie ihn nennen ('**#InternetAsksMerkel** what will or will not be allowed in the Free Trade Agreement and if you had a T-rex what would you name it')

The onomatopoeic representations of laughter that were analysed in this study were limited to a combination of vowels and the letter < h >, such as *höhö* or *ahaha*, as in Examples 7.20 and 7.21. Although the *NottDeuYTSch* corpus contains multilingual comments, so that variants of laughter from other languages outside of German and English are found, such as *jajaja*, *kkkk*, and *555*, these were not ultimately included in the analysis because not all multilingual variants could be captured through RegEx patterns. For example, a popular way of expressing laughter by young Turkish speakers that could not be captured through RegEx is the seemingly random pressing of letters, called *random atmak/random gülmek* ('random throwing/random laughing') (Urhan Torun 2018: 629), e.g. *shahskhhskhsa* (Çelikten & Çelikten 2020: 191). Despite the claim by McCulloch (2019: 6–7) that "keysmashing" or "keyboard mashing" in English has a distinct pattern (e.g. "almost always begins with < a >"), this was not true of

the examples that I was able to identify in the *NottDeuYTSch* corpus.[2] However, there are also functional differences between *random atmak* and *keysmashing*: the former exclusively represents laughter, whereas the latter has been said more broadly to "signal a feeling so intense that you can't possibly type real words" (McCulloch 2019: 6-7).

**Example 7.20** (2013)

[YOUTUBER]. penis höhöhö ich hab penis vesagt ('[YOUTUBER]. penis höhöhö I said penis')

**Example 7.21** (2016)

ich liebe deinr pranks ahaha ('I love your pranks ahaha')

In addition to keysmashing, the use of < kek/kekw > and < sksksk > to represent laughter could not be included in the analysis. < Kek > has increased in frequency on *YouTube* (see Example 7.22), having originally begun as a deliberate misspelling of <lol> on the 4chan message boards (Papasavva et al. 2020).[3] However, in the *NottDeuYTSch* corpus, < kek > is mostly used as an insult, following the popularisation of the word *kek* (originally Turkish for a plain cake) in the Berlin rap scene in the 2010s (see Example 7.23).[4] For this reason, < kek > is not included in the analysis. Similarly, < sksksk > was not included. It was originally used by Black Americans on Twitter as an onomatopoeic representation of laughter and then adopted by "VSCO girls", a term referring to "largely white and affluent" teenage girls (Strapagiel 2019). However, it was not possible to establish if comments containing these patterns in the *NottDeuYTSch* corpus were used to represent laughter or were examples of keysmashing instead.

**Example 7.22** (2018)

[...] Hier ein Glöcklein, hier ein Schimmel, Und am Kuchenstand der kleine

[2]The 'a' key is found in the same position in both the standard QWERTY English-language layout and the QWERTZ German layout, i.e. the first letter of the second line. This rules out potential orthographic differences in keysmashing due to keyboard layout, such as may be the case with the French standard AZERTY layout, where the 'a' key is the first letter of the first line.

[3]*Lol* was first altered to *lel*, which in turn became *kek* due to the proximity of *k* and *l* in the English-language QWERTY keyboard layout.

[4]See Bushido's 2015 album, "Carlo Cokxxx Nutten 3", for several tracks where *kek* is used. The earliest example of *kek* as an insult I have found comes from "Cordon Sport Massenmord", a 2002 track by Sonny Black and Frank White (pseudonyms for Bushido and Fler).

Schlingel. hehe kek :D ('[…] Here a little bell, here mould, And on the cake stand the little rascal. Hehe kek :D')

**Example 7.23** (2017)

[COMMENTER] du kek hast keine ahnung. das is kein rap. das is ist sprechgesang… er sagte selber mal dass er kein rappen is… ('[COMMENTER] you kek you have no idea. that is not rap. that is sprechgesang [spoken singing]… he said himself that he is not a rapper…')

Literal representations of visual metacommunication, such as < icks de > were captured using RegEx patterns that were iteratively developed to account for the many orthographic variations. There are few shared orthographic characteristics of these features, which makes capturing them particularly time-consuming, as they have to be first visually identified, before a RegEx pattern can be developed. For this reason, there may be literal representations of visual metacommunication in the *NottDeuYTSch* corpus that were not captured.

Table 7.3 summarizes the twelve groups of orthographic features that have been identified from previous research and from the examination of the *NottDeuYTSch* corpus and that may fulfil metacommunicative functions. The features form the basis of the creation of the framework of metacommunicative functions, developed in Section 7.4.

Table 7.3: Orthographic features in Digital Writing that fulfil metacommunicative functions

| Orthographic feature category | | |
|---|---|---|
| **Graphicons** | **Creative orthography** | **Metalinguistic signs** |
| ASCII art | Grapheme repetition | Hashtags |
| Emoji | Letter case | Inflectives |
| Emoticons | Punctuation repetition | Laughter |
| Kaomoji | Spaces and line breaks | Literal representation of visual metacommunication |

## 7.3 Analysing orthographic features in the *NottDeuYTSch* corpus

The wide variety of orthographic features identified in the *NottDeuYTSch* corpus that alter the interpretation of a digital text demonstrate the creativity and innovation in Digital Writing for metacommunicative meaning-making. While several features, such as the repetition of graphemes and punctuation, have existed since the 1980s and are still in wide use, as shown in Section 7.1, the data in the *NottDeuYTSch* corpus also include examples of several relatively lesser-used features. These features can be can be grouped together, meaning that, while individually relatively rare, within a groups they can be included in the analysis of metacommunication. This is particularly pertinent for the various literal representations of visual metacommunication, e.g. the use of < icks > to refer to an emoticon, or writing < 1 > embedded within exclamation marks for irony.

The identification process has been aided by the extensive use of RegEx patterns to easily capture orthographically similar features, as we have seen in Section 7.2. In particular, the five-stage Regular Expression created to capture emoticons (discussed in Section 7.2.1) was refined over two years using the corpus data and will be invaluable for future research on data involving emoticons. Furthermore, the large size of the *NottDeuYTSch* corpus has enabled me to identify the emergent use of metacommunicative orthographic features that have originated on other websites but are spreading throughout digital youth language, e.g. the use of < kappa > to demonstrate irony, which was first used on the streaming site *Twitch*, but is now used on *YouTube* (see Section 7.1.3).

Section 7.4 now turns to analysis of the metacommunicative functions of the twelve groups of orthographic features discussed here, presenting a holistic framework of metacommunicative functions in Digital Writing, that incorporates previous work from the fields of DMC, pragmatics, and visual studies.

## 7.4 Developing a holistic framework of metacommunicative functions in Digital Writing

This Section presents a new unified conceptual framework to capture the metacommunicative functions of orthographic features of Digital Writing identified in Section 7, drawing on previous scholarship on metacommunication in DMC, as well as on concepts and frameworks from pragmatics and visual studies. The framework categorises metacommunicative functions into four broad kinds of function:

1. Expressing illocutionary force (7.4.1)
2. Indexing the identity or identities of the author of the digital text (7.4.2)
3. Illustrative effect (7.4.3)
4. Structuring the information within the digital space (7.4.4)

We shall see that this framework of metacommunicative functions can usefully be applied to analyse the orthographic features of digital youth language, and, I argue, for Digital Writing more generally. In addition, the chapter sheds new light on the creativity and innovation in Digital Writing which enables information to be encoded outside of lexical choice within a message.

Over the past 30 years, several frameworks have been developed to analyse the functions of orthographic features in DMC, which I have discussed in Chapter 2, Section 2.1.1.1. However, whilst there have been references to the use of these features to impart metacommunicative functions, the metacommunicative aspect has not been the central focus of analysis. While other frameworks have focused on potential metacommunicative functions, the analyses have been restricted to individual or a small set of orthographic features (e.g. only emoji in Beißwenger & Pappert 2019a). This section re-examines these previous frameworks, within the context of the metacommunicative innovations identified in the *NottDeuYTSch* corpus to develop a new unified framework that is intended to be applicable to all orthographic features used for metacommunicative functions in DMC.

Early frameworks developed to examine digital orthographic features were

primarily used to analyse interactive aspects of DMC, but they did not fully encompass the wide range of metacommunicative functions that were used in digital writing. For example, Collot & Belmore (1996: 15–18) applied the multidimensional-multi-feature model (MD-MF), developed by Biber (1988), to DMC, to examine discourse and interaction in digital "speech situations". However, in terms of metacommunicative functions, the model was restricted to how participants' lexical choice and non-standard spelling were used to establish relations to, and evaluations of, other participants. German-language research, in particular, has primarily examined the relationship between spoken language and many non-standard linguistic features identified in DMC, centred around the concept of conceptual orality (see Chapter 2, Section 2.1.1) and "cross-modal influence" (Baron 1984: 123-124), i.e. a "blurring of traditional distinctions between spoken and written style", that have been considered to be particularly present in DMC (see Storrer 2001; Schlobinski 2005; Wirth 2005; Siebenhaar 2006a). However, this approach does not account for linguistic strategies in colloquial DMC that are not representations of spoken language, such as the use of non-standard orthography to reduce the number of characters used ('language economisation', see Chapter 2, Section 2.1.1.1). While some research has treated all non-standard orthographic features as conceptually oral (Kilian 2001; Bader 2002; Kleinberger Günther & Spiegel 2006), Androutsopoulos (2011: 149) distinguishes three categories of orthographic feature in DMC (as mentioned in Chapter 2, Section 2.1.1.1):

1. "Orality", defined as "all aspects reminiscent of casual spoken language", which includes mostly lexical features, such as the use of slang or interjections.
2. "Compensation", defined as the emulation of "facial expressions or intonation patterns", such as the use of prosodic spelling or emoticons.
3. "Language economisation" (also see Schlobinski 2006b; Siever 2006), i.e. the use of orthographic features to "shorten a message form" to compensate for technological and financial barriers. (Androutsopoulos 2011: 149). This has variously been called "netspeak" (Crystal 2004) or "textspeak" (Crystal 2008), referring to the extensive use of acronyms,

e.g. < hdgdl > for '[ich] habe dich ganz doll lieb', or phonetic spellings, e.g. 'cul8r' for 'see you later', which were popular around the turn of the millennium.

However, these categories are used to more generally analyse DMC and still explicitly refer to a relationship between written and spoken language. The framework developed in this chapter builds on the categories in Androutsopoulos (2011), as well as undertaking other qualitative analyses of previous research to classify the orthographic features in Digital Writing into four broad metacommunicative functions:

1. Illocutionary force
2. Indexing identity
3. Illustration
4. Structuring the information within the digital space

Below, I explain each function in turn with reference to the relevant literature.

### 7.4.1 Illocutionary force

The concept of illocutionary force is a part of the linguistic framework of illocutionary acts, i.e. the unspoken implications of a speech act beyond that literally expressed, where illocutionary force is the strength of the implications delivered by the illocutionary act (Searle 1976: 2-3). Example 7.24 contains two messages, which have "the same illocutionary point" (Searle 1976: 5), i.e. they commit both a directive and commissive illocutionary act by attempting to cause the speaker and listener to visit the cinema,[5] "but [are] presented with different strengths". Here, the difference in verb choice (*suggest* versus *insist*) conveys different implications of potential negotiability of the illocutionary act. Illocutionary acts where less negotiability is offered, or where high emotion is conveyed, are referred to as having a stronger illocutionary force.

**Example 7.24** (Searle (1976: 5))
"I suggest we go to the movies" (low strength)

[5]For more on the different kinds of illocutionary act, see Chapter 2, Section 2.2.4.

"I insist that we go to the movies" (high strength)

While there has been criticism of Searle's approach to illocutionary acts (see Burkhardt 1990; Nastri, Pena, & Hancock 2006), the concept of the illocutionary force of a speech act continues to be used in linguistic and discourse analyses, and is employed similarly in this framework. Several studies of spoken language have examined how lexical items can modify the force of a message (e.g. modal particles in German, see Bross 2012), but the potential of non-lexical features to convey illocutionary force has also been noted. For example, Holmes (1984) lists volume, stress, and pitch as metacommunicative devices that can modify illocutionary force in spoken language, alongside lexical rhetorical devices such as tag questions and rhetorical questions. It is reasonable to expect, then, that techniques in digital writing that may have been originally developed to represent certain features of spoken language can also be used to modify illocutionary force, e.g. the creative use of letter case may represent the illocutionary forces associated with volume, and grapheme and punctuation repetition may represent stress and pitch modulation to alter the illocutionary force of a digital message. Evidence for this presumption in DMC can be found in the study by Nastri, Pena, & Hancock (2006), which showed that non-standard orthography, such as grapheme repetition, can modify the illocutionary force of a message (although lexical choice was the focus of the study). In more recent DMC research, many studies directly reference illocutionary force as a metacommunicative function of orthographic features, particularly emoji and emoticons (e.g. Dresner & Herring 2010; Albert 2015; Siebenhaar 2018; Beißwenger & Pappert 2020) or otherwise refer to how the features accompanying a message signal emotion (e.g. Ahn, Park, & Han 2011; Storrer 2013; Skovholt, Grønning, & Kankaanranta 2014), sentiment (e.g. Novak et al. 2015), or emphasis (e.g. Hilte, Vandekerckhove, & Daelemans 2019), all of which fall under the premise of illocutionary force, as per Searle (1976).

Examples 7.25 to 7.27, taken from the *NottDeuYTSch* corpus, illustrate how the illocutionary force of a message can be modified through the repetition of graphemes and punctuation, and the contrasting use of non-standard letter

case. The repetition of the letter < o > in Example 7.25 can be interpreted as further intensifying the intensifier, *so*. Examples 7.26 and 7.27 demonstrate that the same word or section of a message can have several metacommunicative functions. Although they both contain examples of creative letter case and repeated punctuation, the illocutionary force of the two messages differs substantially. In Example 7.26, the use of lower case lessens the force of the request in the first half of the comment, in contrast with the use of capital letters and repeated exclamation marks at the end of the message, which serve as intensifiers, similar to grapheme repetition.[6] By combining the repetition of punctuation marks and capital letters, the illocutionary force is further increased. However, in Example 7.27 the alternating letter case and the numerals used instead of exclamation marks serve to mark irony, as we saw in Section 7.1.2.

**Example 7.25** (Illocutionary force through grapheme repetitio, 2010)
Das Rumgemieze von [YOUTUBER] am Ende nervt stark :O Aber der rest is **sooo** geil :D mehr von kranker scheiße bitte ;)
('The mewing around by [YOUTUBER] at the end is really annoying :O But the rest is sooo cool :D more sick shit please ;)')

**Example 7.26** (Illocutionary force through punctuation repetition, 2010)
wie heißt das lied bei 1:03 BITTE UM ANTWORT**!!!**
('what is the name of the song at 1:03 PLEASE ANSWER**!!!**')

**Example 7.27** (Illocutionary force through creative letter case, 2011)
Ich bin sooo Cool ich verdünne wasser mit **WAsSeR**
('I am sooo Cool I dilute water with **WAtEr**')

Utz (2000) and Riva (2002: 586) suggested that digital communicators could modify the illocutionary force of their messages through the use of emoticons, and "new emotional tools – emotes and social verbs" (this sense of emotes

[6]Other techniques in Digital Writing that are used to lessen illocutionary force include the use of superscript, e.g. this message, although the *YouTube* comments section does not provide support for such a feature, therefore requiring the commenter to copy and paste the stylised text from a third party website or application.

should not be confused with *Twitch* emotes, examined in Section 7.1.3). Specific to Multi-User Domains (MUDs), text-based digital role-playing environments popular in the 1980s and 1990s (Dieterle 2009), MUD emotes and social verbs are features where users type a short code to make their character 'show' emotion (often just a textual description), e.g. typing "smiiro" would make the user's character produce a text line to "smile ironically" (Riva 2002: 587). While social verbs and emotes of this kind are specific to MUDs, so are not analysed here, the use of emoticons and, later, emoji as illocutionary force markers subsequently became a central focus for DMC research on semiotics and pragmatics. Skovholt, Grønning, & Kankaanranta (2014: 780) argued that rather than "provid[ing] information about how an utterance is supposed to be interpreted [...] emoticons function as contextualization cues, which serve to organize interpersonal relations in written interaction". For Dresner & Herring (2010: 255) "emoticons are used not as signs of emotion, but rather as an indication of the kind of illocutionary force of the textual utterances that they accompany". They reason that in face-to-face communication, people would not necessarily adopt the facial expressions equivalent to the emoticons they were using, as shown in Example 7.28. In Example 7.28a, the use of a winking emoticon mitigates a potentially "face-threatening speech act" (also see Example 7.25), but a wink in a face-to-face situation could be construed as inappropriate. In Example 7.28b, the authors observe that the smiley face "alters the pragmatic meaning of the utterance [...] rather than being a rude, selfish gripe [i.e. the possible interpretation were it not to be followed by a smiley], it becomes a mild, humorous complaint".

**Example 7.28** (from Dresner and Herring (2010: 257-258))

a: I would like a noncircumventing solution ;->

b: JKingsbury: GUIDE> have you ever made a home page on aol?
Guide ASH: JK, yes and I can't get rid of the stupid thing! :)

Similarly, Albert (2015), Hougaard & Rathje (2018), and Weissman & Tanner (2018) all demonstrated that emoji can also be used to modify the illocutionary force of a message, beyond simply depicting facial expressions, despite the

prominence of such emoji in DMC (Emojipedia 2021), e.g. the 'Face with Tears of Joy' emoji (😂). One example is the use of emoji to convey irony in digital writing. Building on previous work on emoticons by Thompson & Filik (2016), Weissman & Tanner (2018) showed how the use of an emoji that is perceived to be "mismatched" with the sentiment of the message (e.g. a negative message paired with a smiley face) is reliably perceived as an indicator of irony by readers. Furthermore, the same emoji can be used to modify illocutionary force in different ways. Example 7.29 shows how the middle finger emoji can be used to intensify the anger directed at Bayern Munich, as in (a), or provide a humorous effect when paired with the crying with laughter emoji, as in (b).

**Example 7.29** (The middle finger emoji providing different illocutionary force)
a (2016):
Scheiß Bayern 🖕🖕
('Bloody Bayern [Munich]🖕🖕')

b (2017):
[NAME] du Lauch 😂🖕
('[NAME] you mug [lit. leek] 😂🖕'🟩)

In addition to the examples given above, the illocutionary force of a message can be altered through the manipulation of spacing and metalinguistic signs, something which has almost entirely escaped the attention of previous research (though for an investigation of the use of < kappa > on *Twitch*, see Barbieri et al. (2017)). In Example 7.30, the insertion of spaces between each character can be interpreted as an exaggeration of the author's disbelief, perhaps alluding to more confusion, compared what would be signalled were the message just written in capital letters ("WHAT?"). Example 7.31 shows the use of the gaming-specific metalinguistic sign, < kappa >, to signify irony or a joking demeanour, which is then further reinforced by writing < xd >, the lowercase variant of the < XD > emoticon, which resembles a face with eyes creased laughing.

**Example 7.30** (Illocutionary force through letter spacing; 2010)

W H A T ? Ich bin nicht nur geschockt sondern auch schockiert. xD
('W H A T ? I am not only shocked but also in shock. xD')

**Example 7.31** (Illocutionary force through metalinguistic signs; 2016)
10000 deabos #Veränderung Kappa xd
('10000 unsubscribes #Change Kappa xd')

The Examples 7.25 to 7.31 in this section thus demonstrate that each of the groups of orthographic features identified in Section 7.1 can be used to modify the illocutionary force of the message, and that this is not true only of graphicons, which have been the primary focus of previous research. Illocutionary force also encompasses several related concepts that have been used to describe possible metacommunicative functions of orthographic features, including emotion, stress, and sentiment. For these reasons, illocutionary force is an appropriate choice of function through which to analyse metacommunication.

### 7.4.2 Indexing identity

A second metacommunicative function of orthographic features is to index particular sociolinguistic meanings within the comment. The concept of indexicality, developed by Silverstein (1976, 1993, 2003), has been defined as

> the dimension of meaning in which textual features 'point to' (index) contextually retrievable meanings. More concretely: every utterance carries apart from 'pure' (denotational) meanings a range of sociocultural meanings, derived from widespread assumptions about the meanings [by author and reader]
> (Blommaert, Westinen, & Leppänen 2015: 122)

These sociocultural meanings intertwined with the message can be a product of several sociological identities, including ethnicity (e.g. Le Page & Tabouret-Keller 1985; Nakamura 2002; García 2010; Omoniyi 2016); gender (e.g. Herring 1996c; Huffaker & Calvert 2005; Butler 2006; Herring & Dainas 2020); sexuality (e.g. Pullen & Cooper 2010; Jones 2012; Milani 2013; Borba 2019); and the inextricable intersectional identities between them (e.g. Crenshaw 1989; Clammer

2015; Barbieri & Camacho-Collados 2018); as well as more local identities, such as membership of specific online communities (e.g. Papacharissi 2011; Hamilton, Garretson, & Kerne 2014; Meer 2018).

Digital writing, too, can "index a range of political, cultural or aesthetic orientations that are simultaneously localised within digital culture and linked to global semiotic and cultural flows" (Androutsopoulos 2011: 155-157). Existing DMC scholarship has primarily focused on how lexical choice is used to index identities (e.g. Reershemius & Ziegler 2015; Lee 2016; LaViolette 2017; Røyneland 2018), although more recent studies have started to examine orthographic phenomena, such as punctuation use to index certain youth identities (Androutsopoulos 2018); stylised spelling to index sexual identities (Ilbury 2020); and emoji to index ethnic and racialised identities (Robertson, Magdy, & Goldwater 2018), gender identities (Aull 2019) or an intersectional approach to both gender and ethnicity (Barbieri & Camacho-Collados 2018).

Barbieri & Camacho-Collados (2018) present useful evidence of how the indexical meanings of the same base emoji differ based on their combination with skin tone modifiers or gender modifiers (although the study did not take gender-neutral emoji into account).[7] Examples 7.32 to 7.36 from both the *NottDeuYTSch* corpus and the data from Barbieri & Camacho-Collados (2018: 102–3) illustrate the metacommunicative differences conveyed by the uses of the 'Raised Fist' emoji, a symbol of protest with a long history for both class and race-based rights, with different skin tone modifiers. In Example 7.32, the 'Raised Fist' emoji is unmarked for skin tone and can be interpreted as a demonstration of solidarity with the recipient of the message. Examples 7.33 and 7.34 contain the 'Raised Fist' emoji with medium-dark skin tone and can be interpreted as a demonstration of racial solidarity as they are explicitly marked with a skin-tone modifier (Barbieri & Camacho-Collados 2018: 102): Example 7.33 shows solidarity with Colin Kaepernick, an American football player who protested racial injustices in the United States of America, and Example 7.34 shows support for the 'Young Movement Gang', an internet collective founded

[7] See Chapter Section 7.2.1 for an explanation of the how emoji characters are combined to produce new emoji and emoji variants based on gender and skin tone.

by Black German *YouTuber*, Simon Desue. Despite being explicitly marked for light skin tone, the 'Raised Fist' emoji in Example 7.35 can be interpreted as indexing the ethnic background of the commenter, but the context of the message, i.e. respecting others regardless of nationality, indicates that racialised solidarity between white people is not indexed here, rather it is more likely a demonstration of more general solidarity with all potential readers of the comment. The 'Raised Fist' in Example 7.36, however, despite not being marked for skin tone, can be interpreted as an indication of solidarity between members of the same minoritised ethnic group, due to the explicit divulging of ethnic identity, *Bin Auch Türkin* ('Am Also Turkish'). The unmarked emoji in this example could also be interpreted as the demonstration of solidarity based on shared membership of a *YouTube* community or interest in food (due to the mention of baklava), rather than necessarily as explicitly race- or ethnicity-based; alternatively, it is possible that the device used to send the message did not have the capability to customise the skin tone of the emoji, as this was only added in November 2015 (Emojipedia 2017).

**Example 7.32** (2016)
die machen wir fertig ✊
('We'll sort them out ✊')

**Example 7.33** (adapted from Barbieri and Camacho-Collados (2018: 102))
#Kaepernick we riding with you bro ✊🏾

**Example 7.34** (2017)
YMG ✊🏾

**Example 7.35** (2018)
Nationalität ist egal solange man Stabil ist und andere Menschen respektiert ✊🏻
('Nationality is not important as long as you are cool and respect other people ✊🏻')

**Example 7.36** (2018)

[COMMENTER] Jaaaaa [YOUTUBER] Isst Garnichts 😻😻 P.S Bin Auch Türkin Und Ich LIEBE Baklava 🌈🌈😻😻✊✊
('[COMMENTER] Yeaaaah [YOUTUBER] Eats Nothing 😻😻 P.S Am Also Turkish And I LOVE Baklava 🌈🌈😻😻✊✊')

In the *NottDeuYTSch* corpus, orthographic features can be used to index affiliation not only with broader ethnic and gender identities, but also with specific online communities or even particular channels. For example, the use of < kappa > indexes affinity to the online gaming community, having originated on the video game streaming website *Twitch*, as examined in Section 7.1.3. In the *NottDeuYTSch* corpus, the usage of < kappa > occurs in over 87% of videos (27 out of 31) where the *YouTuber* produces gaming content, which reflects the overlap in interest with *Twitch*. From a theoretical perspective, in using < kappa >, the comment author is committing an "act of identity" (Le Page & Tabouret-Keller 1985), which indexes their status as belonging to the "in-group" of online gaming (see Tajfel & Turner 1979), familiarity and potential "in-group" status among the community, as in Example 7.37, where commenters use < kappa > in their discussion of a *YouTuber* spending large sums of money on a gaming PC.

**Example 7.37** (2018)
[IN REPLY TO COMMENTER] und 30 k wegwerfen sind ja nix neh 😂 (**kappa**)
('[IN REPLY TO COMMENTER] and throwing away 30 k is nothing eh 😂 (**kappa**)')

Orthographic features can index identities with far smaller membership than gender or ethnic identities or popular shared interests such as gaming. Example 7.38 contains a string of three 'Unicorn' emoji at the end of the comment, which might normally be interpreted as an intensification of a positive communicative tone, similar to the use of the 'Smiling Face with Heart-Eyes' emoji earlier in the comment.[8] However, within the German *YouTube* scene, the 'Uni-

[8]Of the 99,495 comments in the *NottDeuYTSch* corpus containing 'Unicorn' emoji, 99% had a positive emotional sentiment, according to an analysis using the SentiWS dataset of German

corn' emoji has been co-opted by *YouTuber* BonnyTrash as part of her channel's brand identity, incorporating it within her logo (see Figure 7.7) and merchandise (e.g. hoodies with an 'All I want for Christmas is a unicorn' slogan, see her official store Bonnytrash 2018). While unicorn imagery has been heavily marketed towards children in recent times (Weida & Bradbury 2020), as well as becoming an LGBT+ symbol (Balirano 2020), qualitative and statistical analyses of its use in the *NottDeuYTSch* corpus point to the desire of commenters to index their affiliation to BonnyTrash as the reason for the high frequency of usage in the comments under her videos, rather than indexing other identities, such as being a child or considering oneself a member of the LGBT+ community. Of the 99,495 'Unicorn' emoji in the *NottDeuYTSch* corpus, 59% (n = 58,655) were found in comments under BonnyTrash's videos. This means that 23% of all the emoji appearing under her videos are 'Unicorn' emoji, around nine times greater than the frequency of 'Unicorn' emoji in the whole corpus, a very significant difference, $\chi^2$ (1, n = 1,565,549) = 100432.8553, p < 0.001). Considering the lack of overt LGBT+ content in BonnyTrash's videos, using the 'Unicorn' emoji in a comment under her videos can therefore be interpreted as commenters indexing an identity as a member of the channel's community, functioning as a marker of in-group status.

**Example 7.38** (2018)
Hey ich liebe 😍3 Uhr nachts wiedios 🦄🦄🦄💁
('hey I love 😍3am videos 🦄🦄🦄💁')

### 7.4.3 Illustration

A third metacommunicative function of orthographic features in DMC is to provide information relevant to the message through the use of graphically distinct content that adds to, refers to, or replaces lexical choices to provide context for the reader. I call this illustration. For example, the use of the 'Two Men Holding Hands' emoji in Example 7.39 directly follows the word "brothers", representing

language (Remus, Quasthoff, & Heyer 2010), compared to 60% of other comments containing emoji (p < 0.0001).

Figure 7.7: BonnyTrash homepage banner

and reinforcing the reference to the close relationship of the two men.[9] In face-to-face communication, metacommunicative illustration occurs through body movements or actions, such as approximating shapes or sizes, mimetic (acting, imitating), and deictic (pointing) gestures. In digital writing, a wide range of orthographic features can be used for mimesis and deixis as part of illustrative metacommunication, chiefly graphicons and other creative combinations of ASCII characters. The frequent use of graphicons is partly due to the availability since 2010 of over 2,000 emoji in the 'People & Body' group, as body parts, gestures, activities, and roles can now be depicted (e.g. < 🤾 >, 'Woman Playing Handball'), not just the range of facial expressions (e.g. < 😮 >, 'Face with Open Mouth').

**Example 7.39** (2015)
Irgendwie süß ihr beiden - wie Brüder👬
(Kind of sweet you two - like brothers👬)

While all metacommunicative functions can be analysed using relevance theory, as the orthographic features are used for the most part to "guide the hearer towards the speaker's meaning" (Wilson & Sperber 2002: 607), relevance is particularly central to the metacommunicative function of illustration. In particular, metacommunicative illustration conveys a "positive cognitive effect" to the reader, i.e. an input that helps provide the correct context through reference

[9] An additional interpretation of the use of "brothers" could be an ironic reference to the trope of casual LGBT+ erasure in media and other spaces (see Waggoner 2018), as the 'Two Men Holding Hands' emoji is popularly used to represent a gay couple.

to "observable phenomena", "thoughts", "memories", and "conclusions of inferences" (Wilson & Sperber 2002: 607). In addition to the visual representation of the message content through the use of emoji in Example 7.39, a deictic gesture constructed of ASCII symbols is shown in Example 7.40, i.e. an arrow pointing left constructed by using a corner bracket or 'less-than sign' followed by a dash (<–). The deictic gesture provides the context, i.e. the entry in a competition symbolised by the hashtag, for the request to be chosen.

**Example 7.40** (2015)
#SamsungVR **<–** BITTE MAMA! BITTE!
('#SamsungVR [virtual reality headset] **<–** PLEASE MUM! PLEASE!')

A central defining characteristic of orthographic features used for metacommunicative illustration is their interpretation by the reader as distinct from the referent. This is primarily achieved through the difference in the type of character between referent and orthographic feature, e.g. a sentence (referent, most likely comprised of alphanumeric characters) and emoji (feature). Characters are generally (but not strictly) interpreted as a different type if they fall into either different Unicode character categories (e.g. letters, numbers, punctuation, symbols, see The Unicode Consortium 2021c: section 5.7.1) or into a different block range (The Unicode Consortium 2021a).[10] In other cases, the same character category could be used but the illustrative function may be signalled in other ways, such as by bracketing, for example the decorative use of < I > in *IIIII EXAMPLE IIIII*, although there are no such cases in the *NottDeuYTSch* corpus. Other plausible distinctions are various kinds of spatial distinction, such as the separation of referent and orthographic feature through the use of spaces, line breaks or other separator characters (or characters functioning as separators, e.g. repeated full stops).

Examples 7.41 to 7.44 demonstrate the mimetic and deictic illustrative functions found in the *NottDeuYTSch* corpus. Deictic gestures, as mentioned before, are

[10] A block range is a set of contiguous character codes, e.g. U+00F6 is the character code for < ö > (referred to as 'Latin Small Letter O with Diaeresis'). The characters in each block range are typically related to each other, so < ö > belongs to the 'Latin-1 Supplement' block between < õ > and < ø >.

not restricted to depictions of arrows or pointing fingers. They can also be metaphorical or abstract, representing what the speaker could be indicating towards, shared physical experiences, or hypothetical visual imagery. Example 7.41 uses the obvious connections between an angler and their intended catch, a fish, to represent the content of the message through the use the combination of the 'Fishing Pole and Fish' and 'Fish' emoji. Example 7.42, from Dürscheid & Siever (2017), is more metaphorical. Dürscheid & Siever (2017), when investigating the role of emoji as punctuation, speculated on the function of the 'Sun' emoji following an exclamation mark, positing that it could represent either a whole sentence, i.e. "the sun is shining", or that is was an extension of the greeting, *Guets mörgeli* ('Good morning'). I would argue that the 'Sun' emoji adds contextual information to the greeting, representing the physical or sensory experiences related to morning sunshine, which can include rays of light breaking through the curtains, or the warmth of the sun on one's skin. Similar effects are achieved through other weather or season-related emoji, such as the 'Snowman' emoji in Example 7.43, which, for example, suggest possible actions, e.g. building a snowman, in such weather, or the 'Christmas Tree' emoji in Example 7.44, which reminds readers of the cultural and sensory experiences relevant to a decorated fir tree.

**Example 7.41** (2016)
[YOUTUBER] ist Anglerin.🎧🐟
('[YOUTUBER] is an angler.🎧🐟')

**Example 7.42** (adapted from Dürscheid and Siever (2017: 281))
Guets mörgeli! 🌞 scho erwachet?
('Good morning! 🌞 already awake?')

**Example 7.43** (2017)
bei uns ist es gefühlte 6- eiskalt ☃❄
('around us it feels like 6- ice cold ☃❄')

**Example 7.44** (2017)
Frohe Weihnachten 🎄

('Merry Christmas 🎄')

I would also contend that the use of graphicons to represent facial expressions can have an illustrative function. Example 7.45 uses the 'Face with Open Mouth' emoji to mimetically illustrate the face the commenter might plausibly make whilst viewing the video and writing the comment. These graphicons can also impart illocutionary force, especially when repeated, such as in Example 7.46.

**Example 7.45** (2017)
Ich erkenne ihn garnicht mehr😮
('I don't recognise him any more 😮')

**Example 7.46** (2017)
Wieso bist du nackt 😮😮😮😮😮😮
('Why are you naked 😮😮😮😮😮😮')

Graphicons, in particular, can be used to impart mimetic illustrative functions by acting as a visual representation of an object, in place of, rather than alongside, the word(s). In Example 7.47, Na'aman, Provenza, & Montoya (2017: 137) provide an instance where the 'Key' and 'Pizza' emoji are embedded within the sentences, replacing the words. Dürscheid & Siever (2017), who have examined this function in Swiss digital messages, noted that emoji can also be used as allographs, replacing individual letters, not just whole words. In Example 7.48, the < o > in *Bonne Nuit* is replaced by the 'First Quarter Moon Face' emoji, which is thematically linked to the message.

**Example 7.47** (adapted from Na'aman et al. (2017: 137))
The🔑 to success is 🍕
('The key to success is pizza')

**Example 7.48** (adapted from Dürscheid and Siever (2017: 279))
B🌛nne nuit !
('G🌛🌛d night !')

The allographic replacement of words or letters by graphicons provides little ex-

tra information about how the message should be received but, in some cases, could lead to a quicker positive cognitive effect of the entire message by lowering "the processing effort expended" by the reader (Wilson & Sperber 2002: 609), although perhaps there is slower processing of the words altered by allographic replacement. Indeed, the change in cognitive processing speed has been shown in psycholinguistic analyses of both emoticons (e.g. Thompson & Filik 2016) and emoji (e.g. Holtgraves & Robinson 2020). Allographic replacement can occur in more stylised forms and does not even need to be embedded within text. Example 7.49 shows the first line of an emoji 'translation' of Moby Dick, where we see the 'Telephone' emoji used metonymically for the verb "to call" (Emoji Dick, Benenson 2010, the emoji text and the original English appear on facing pages). Nor does the metacommunicatively illustrative use of orthographic features need to be embedded within text. Furthermore, in *YouTube* comments, the referent in illustrative metacommunication may be the video, rather than other elements of a comment. Examples 7.50 and 7.51, taken from the *NottDeuYTSch* corpus, contain two comments where strings of emoji (and non-alphabetical characters) convey the message instead of words. The repeating string of the 'Father Christmas', 'Christmas Tree', and 'Present' emoji in Examples 7.50 metacommunicatively illustrates the Christmas theme of the video content under which the emoji are written. In Example 7.51, the 'Boy: Medium Skin Tone', 'Girl: Medium Skin Tone', and 'Boy: Medium-Light Skin Tone' emoji represent the participants in the video, the 'Raised Hand' emoji refers to the central theme of 'facepalming' in the video, a gesture of placing a hand on one's face to indicate frustration or embarrassment. Finally, the 'Volcano' emoji can be interpreted as a positive evaluation of the video, i.e. the video is exploding or rapidly gaining recognition.

**Example 7.49** (adapted from Benenson (2010: 15))

☎👦⛵👍👌

('Call me Ishmael')

**Example 7.50** (2012)

🎅🎄🎁🎅🎄🎁🎅🎄🎁🎅🎄🎁

**Example 7.51** (2016)

✋+🧒+👧+😇=🙏

As can be seen in these examples, there is often a ludic element to the use of features for illustration, which reinforces the informal element of digital vernaculars (Androutsopoulos 2011: 155), such as communicating emotional proximity between speakers (Kelly & Watts 2015). Many features, not just emoji, are used ludically in the *NottDeuYTSch* corpus, as might be expected of vernacular communication between young people consuming entertainment. Example 7.52 and Figure 7.8 show some of the ways in which orthographic features can be used ludically to illustrate the message. In Example 7.52, the multiple thematically similar graphemes and punctuation do not merely intensify or alter the illocutionary force of the message (nor are they supposed to be allographic replacements of lexical elements), they ludically illustrate the message through repetition *ad absurdum*. In this example, the use of multiple 'Smiling Cat Face with Heart-Shaped Eyes' emoji also increases the illocutionary force, i.e. the commenter's appreciation of the video. Figure 7.8 shows a humorous use of allographic replacement. The replacement of the letter < A > by the 'A Button (Blood Type)' emoji is not related to the content of the message (unlike the 'Crescent Moon' emoji in Example 7.48 above), and, on the face of it, provides little extra information. However, this can be interpreted as a deliberate subversion of 'standard' emoji use, with the effect of being both illustrative and potentially indexing a particular identity, i.e. someone who is familiar with the tropes and memes surrounding emoji use in internet culture.[11]

**Example 7.52** (2016)

Meeeeeeggggaaaa!!!😻😻😻💝💝🍃♡💝❁🦄

[11] The replacement of letters with emoji letters, particularly the use of the 'B Button (Blood Type)' emoji (🅱) is an oblique reference to linguistic practices by the Bloods, the US-based street gang, who, in their rivalry with the Crips, would seek to avoid using [k] sounds or the letters < c > (and sometimes < k >) and replace them with [b] sounds and the letter < b >. In online spaces, this has been the subject of parody, spawning spin-off meme content, often used to parody a 'gangsta' identity by young people far removed from West Coast gang-based culture.

Figure 7.8: Ludic illustration using emoji, from Dürscheid and Siever (2017: 270)

### 7.4.4 Structuring the information within the digital space

The potential metacommunicative function of altering how a message appears on-screen was noted above with regard to changing letter case and the repetition of characters. However, as identified in previous scholarship, these metacommunicative functions were originally developed to represent spoken features, e.g. volume and prosody to alter the illocutionary force of a message. A separate aspect of metacommunication, particular to written communication alone, involves the manipulation of how the message occupies the on-screen space to alter how the message is processed and interpreted by the reader, which I refer to as the metacommunicative function of structuring the information within the digital space. This function tends to exploit the use of separation characters, such as line breaks and spaces, as noted in Section 7.1 with some ASCII art, which often deliberately disrupt the expected user experience of online spaces like the *YouTube* comments section. However, as we shall see, the metacommunicative structuring of information within the digital space can be imparted using a wide variety of orthographic features.

The development of the framework for this metacommunicative function in digital writing comes from research that has otherwise been restricted to handwritten or printed messages (e.g. McAteer 1992), visual poetry (e.g. Bohn 2001, 2010; Elleström 2016), and information design (e.g. Jacobson & Jacobson 1999; Pettersson 2002). It will be helpful to outline briefly the *YouTube* video page navigation experience. As a visitor to a *YouTube* video scrolls down the page, they are confronted with information on the video, other videos to watch, and, as they scroll further down, the comments on the video. The user experience varies depending on the device used to browse the video, as screen size or device type determines the number of characters on each line of the comment, which in turn can change when a comment is "cut-off", i.e. how many characters or lines are the first are shown before the visitor must click 'read more' to expand and read the full comment. Orthographic features can be used to deliberately alter this user experience, and the function of structuring the information within the digital space is rooted in the actual cognitive processing of the message and its digital context.

The metacommunication encoded in these features tends to be more nuanced than other functions, for example, using an emoji marked as a man to index gender identity. At its core, the metacommunicative function of structuring the information within the digital space is inherently artistic, and provokes a reaction in the reader through the disruption caused by orthographic features to the expected scrolling experience for the reader and the line-by-line, left-to-right reading experience. This manipulates the space occupied by the message and therefore affects how it is likely interpreted. Figure 7.9 contains a sentence *[NAME] ist soooooooo* followed by a figure posing, accompanied by the letters of the word, *fabulous*, along the right-hand side. If the figure were to be removed, the basic sense of the message would still be conveyed, i.e. the person in question would still be called fabulous. However, the choice to structure the adjective, *fabulous*, in such a way creates a "symbiotic relationship", in the words of Bohn (2010: 14), between the word and the ASCII art, where the word takes on characteristics of the figure, connecting it with associations, in this case, of dancing or posing, for example.

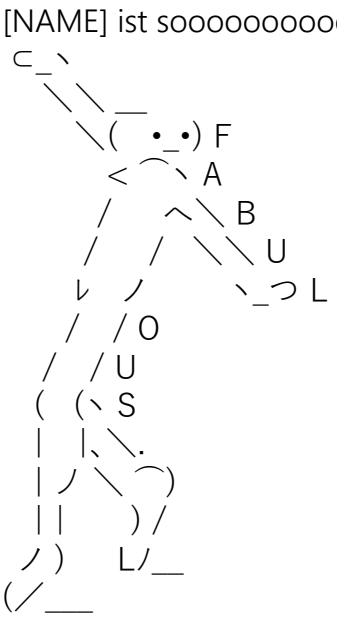



Figure 7.9: ASCII art of a person dancing, 2016

Similarly, Figure 7.10 contains a viral ASCII art piece showing a tank and a man to protest against the policy of *Google* to force *YouTube* users to use *Google+* accounts in order to comment (Blue 2013). The comment is deliberately four lines long and no more than 100 characters wide, in order to occupy the maximum amount of space in a *YouTube* comment without activating the 'read more' function, thus ensuring the maximum amount of visual disruption.

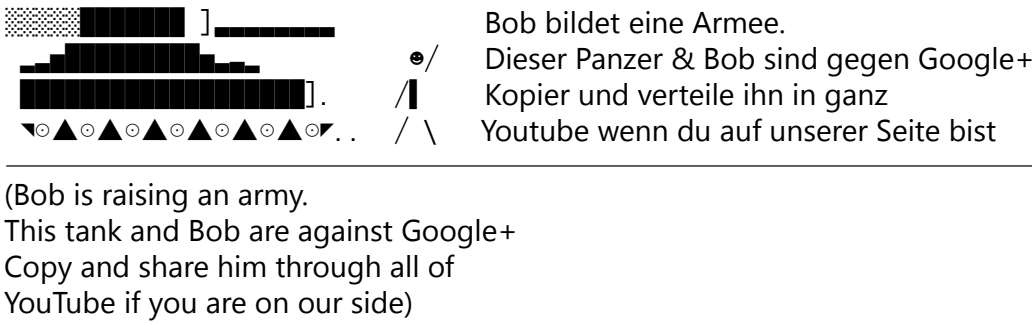



Figure 7.10: ASCII art of a tank and text, 2014

However, the manipulation of text structure does not need to involve ASCII art. The metacommunicative function may be imparted through divergence from expected norms, even on a mundane level. For example, *YouTube* comments (and other sites of digital writing, e.g. *WhatsApp*) are not expected to

have paragraphing, like in standard writing (Schneider 2013: 65), so the use of extensive paragraphing alters what Wyss & Hug (2016) called the "spatio-visual demarcation" of the communicative space and therefore the interpretation of the message. Example 7.53 shows the use of repeated line breaks to separate the questions in a message. The addition of line breaks slows down how the message is read and occupies more space on the page, which could make the reader notice it and cause them to read the message in the first place, and so in turn encourage subsequent interaction. Without such spacing, the comment might instead be interpreted as a more hurried barrage of questions, e.g. *Xbox oder Wii U?blau oder grün?Saft oder Wasser?*. Line breaks can also be used to emphasise certain words, as in Example 7.54, or to deliberately interact with the functionality of the webpage, as shown in Figure 7.11. Here, the author deliberately 'hides' the final sentence by inserting enough line breaks to trigger the automatic truncation of *YouTube* comments, so that the cut-off text is replaced with the 'read more' button. This, again, alters the cognitive processes associated with processing the text.

**Example 7.53** (2014)
Xbox oder Wii U?
blau oder grün?
Saft oder Wasser?
Kunst oder Musik?
Beantwortet ihr mir diese Fragen bitte!??

('Xbox o Wii U?
blue or green?
Juice or water?
Art or music?
Can you answer these questions for me please!??')

**Example 7.54** (2015)
DIGGA.....DU HAST DIE PUNCHLINES GEFLOWT!!!!!!!!!!DAS WAR —>FRESH <—
#Fresh

('MAN.....YOUR FLOW OF THE PUNCHLINES!!!!!!!!!!THAT WAS
—>FRESH <—
#Fresh')

Figure 7.11: *YouTube* comment expanded and with "read more" button

In addition to line breaks, a commenter may alter the interpretation of the message by 'flooding' characters. I use the term 'flooding' here, following Hilte, Vandekerckhove, & Daelemans (2019), to mean an extreme form of repetition, often extending over multiple lines, as demonstrated in Example 7.55 with the repetition of the graphemes in the word *geil*. The repetition here manipulates the space occupied by the message, altering how it is interpreted, according to relevance theory:

> Relevance theory claims that what makes an input worth picking out from the mass of competing stimuli is not just that it is relevant, but that it is more relevant than any alternative input available to us at that time.
> Wilson & Sperber (2002: 609)

In this case, the repetition certainly imparts illocutionary force, intensifying the emotions communicated by the lexical choice, but I would argue that the extreme amount of repetition, causing the message to occupy a far larger amount of space than expected, alters what could be interpreted as the most relevant part of the message and, in doing so, signals that a separate metacommunicative function is also being communicated. In this case, the extreme use

of grapheme repetition may be interpreted as meta-metacommunicative: the commenter shows awareness of the metacommunicative function caused by the repetition of the graphemes (i.e. an increase in the illocutionary force), but also of the fact that this function can be manipulated to change the relationship between the message and the comment space, in this case occupying a far larger portion of the screen.

**Example 7.55** (2013)

ggggggg

👂👂👂👕👕😀
('you poor thing 😷🐄🐄[...]👕👕😀')

**Example 7.57** (2018)
Hey🍂[YOUTUBER]🍂ich🍂 würde🤝 mich😍riesig🍂 auf👑das🎗 Einhorn🐴kuscheltier🐴 freuen🤝 bist🍂dir🐴beste🍂 youtuberin♡ mach🤝 weiter🤝so🎗und🎗 du ♡bist♡ echt♡ hübsch♡ #lysm#Unicorn #verlosung♡♡♡♡♡♡♡♡
('Hey [YOUTUBER] I would be so happy with the unicorn cuddly toy you're the best YouTuber keep on going and you are really cute #lysm[love you so much]#Unicorn#lottery')

## 7.5 Future applications of the metacommunication framework

The four functions outlined above – illocutionary force, indexing identity, illustration, and structuring the information within the digital space – constitute a useful framework of analysis to capture the range of metacommunicative functions carried out by orthographic features in DMC. Existing frameworks of orthographic features in Digital Writing have either not been based on metacommunicative function or have restricted the analysis of metacommunicative functions to smaller groups of features. The framework presented here, in unifying studies of DMC, pragmatics, and visual culture, with my own close analysis of the *NottDeuYTSch* corpus, offers a holistic, flexible framework to analyse metacommunication that can be applied across Digital Writing. The framework allows for the analysis of orthographic features fulfilling multiple metacommunicative functions, as demonstrated in the examples throughout the chapter. A message may contain features that fulfill any or all of the functions simultaneously. Figure 7.12 shows a hypothetical comment, adapted slightly from the *NottDeuYTSch* corpus,[12] where the 'Flexed Biceps: Medium Skin Tone' emoji fulfils all four functions. The message is initially illustrated by a single emoji,

[12]The original comment from the *NottDeuYTSch* corpus reads: "Hahaha das auto ist ja richtig krass [YOUTUBER]💪💪💪💪" ('Hahaha the car is well cool [YOUTUBER]💪💪💪💪').

which can represent an affirmational gesture or body language, analogous to the phrase *starke Leistung* ('strong performance'). The 'Medium Skin Tone' modifier of the emoji indexes a racial or ethnic identity of the comment author, which, in this case, (and similarly to the discussion of the 'Raised Fist' emoji in Section 7.4.2 above), can be interpreted as a demonstration of solidarity with the *YouTuber*, perhaps due to a shared cultural background. The illocutionary force of the message is increased by the repetition of the emoji; both the sentiment conveyed by the phrase *richtig krass*, and any potential ethnic solidarity can be interpreted as being intensified through the use of the emoji. Finally, the extreme repetition of the emoji occupies a significant amount of space on the device screen as the reader scrolls through the comments, potentially drawing more attention to the message, and altering the information design and, therefore, the reading experience and its subsequent interpretation.

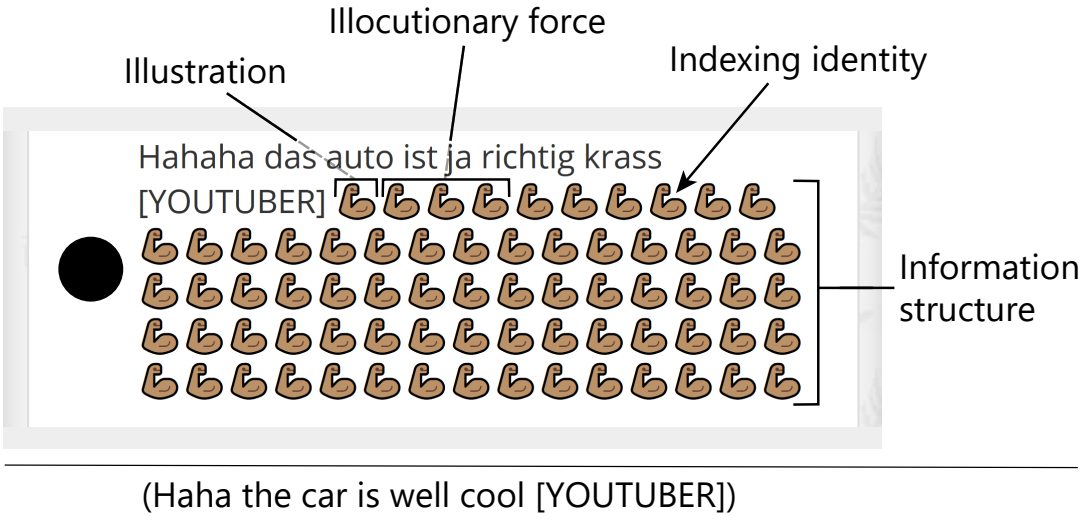



Figure 7.12: The four metacommunicative functions expressed using a single orthographic feature

The framework presented here invites future studies of how orthographic features interact with each other to alter the metacommunicative functions of the message, for example, the relationship between the extent of the repetition of an orthographic feature and the strength of the illocutionary force conveyed in a message, such as the flooding of 'Flexed Biceps' emoji in Example 7.12 above. While the framework has been developed using text-based data, it is

also applicable to other areas of DMC, such as multimodal digital interaction, which is rapidly increasing in prevalence (see Moschini 2016; Barbieri et al. 2018; Herring et al. 2020a). For example, the use of GIFs (i.e. small files that can embed a short video in DMC), is an under-researched area (see Tolins & Samermit 2016), for which the framework could be useful. Furthermore, the framework enables more in-depth analyses of certain orthographic features, such as emoji. I anticipate that emoji ethnography will become a productive sub-field in the near future, as the number of emoji continues to increase, partially fuelled by the drive to include more representative emoji, such as gender-neutral versions and different skin tones for existing gendered emoji (Burge 2019b). The analytical framework of metacommunicative functions presented here may offer a useful basis to explore the nuances of emoji choice, as well as the functions of the wide range of orthographic features used in DMC.

# 8

# Conclusions

This thesis has investigated the linguistic features of digital German youth language through the corpus-linguistic analysis of the *NottDeuYTSch* corpus of *YouTube* comments. The two fields of Digitally Mediated Communication (DMC) and German youth language are well-developed, but this thesis is situated at the intersection of the two, i.e. digital youth language, which has not seen the same level of research. In particular, the thesis has focused on the language of *YouTube* comments written by young people, an area that is, to date, under-researched, despite the popularity of *YouTube* with young people. The *NottDeuYTSch* corpus, constructed especially for this study, spans ten years from 2008 to 2018 and contains over 33 million words and is the first large-scale corpus of YouTube language, impacting future studies of language used by young people, Digital Writing, and media studies. The methods used to create the corpus, as laid out in Chapter 3, including data selection, collection, and cleaning, may facilitate future corpus construction containing *YouTube* comments, as well as more generalised Digital Writing. Indeed, one of the key research outcomes of the thesis was the construction of the *NottDeuYTSch* corpus.

The corpus significantly contributes to the available corpus linguistic resource pool in this field, namely providing a specialised corpus of online language written by young people. The size of the corpus can facilitate future research of emergent linguistic features and innovations, only some of which have been discussed in the thesis. Furthermore, the span of the *NottDeuYTSch* corpus contributes to how we can view linguistic change, especially amongst young people. In the present study, the corpus has enabled quantitative and longitudinal analyses of key questions that had, by necessity, previously been chiefly analysed using qualitative methods. Furthermore, the recent integration of the *NottDeuYTSch* corpus into the *DeReKo* enables similar research to be carried out by future researchers and demonstrates the contribution of the thesis to the field.

Using the *NottDeuYTSch* corpus, the study has provided significant evidence for emergent linguistic innovations and longitudinal changes in digital youth language in lexis and morphology; in syntax, and in Digital Writing, thus addressing the overarching research question of this study: what can we learn from a corpus-driven, longitudinal approach to digital youth language?

Chapter 4 carried out one of the first in-depth lexical and morphological studies of *geil*, despite the ubiquity of the word in both academic studies and media reportage on youth language. Specifically, the chapter demonstrated that *geil* is used in German youth language in three different senses: 'sexually aroused'; 'desperate for something'; and 'cool, great', and that it can be used as various grammatical categories, including as an adjective, adverb, and discourse marker. Using a primarily quantitative methodology, Chapter 4 showed that the morphological status of *geil* within a construction is linked to the sense, with usages of *geil* in the sense to mean 'desperate for something' occurring when *geil* functions as an affixoid, whereas in constructions where *geil* is the base lexeme, the sense tended to be either 'cool, great' or 'sexually aroused'. Furthermore, the chapter identified that constructions where *geil* occurs as an affixoid are likely to be semantically related to *YouTube*, as shown in the use of constructions such as *likegeil* ('desperate for "likes" '), which is linked to the online practice of clicking the 'like' button underneath a *YouTube* video (or other

digital content) as an act of agreement or support. These findings indicate the importance of taking the digital platform into account when analysing language, and one might expect different lexical and morphological innovations on different platforms, such as messaging services (e.g. WhatsApp or Snapchat).

Chapter 4 offered the first in-depth quantitative morphological study of *geil*, in particular analysing the use of intensifiers when *geil* functions as the base lexeme. The chapter demonstrated the wide range of intensifiers that are used with *geil*, and the innovative use of multiple intensifiers in a single *geil* construction. The findings contribute to previous studies of intensification in German, offering a perspective from Digitally-Mediated Communication that not only demonstrates the creativity of young people's Digital Writing, but also expands our understanding of morphological intensification (alongside orthographic intensification, as discussed in Chapter 7). Indeed, there is much material for future research in this area, such as the link between morphological or orthographical complexity and degree of intensification. Furthermore, using a mixed methods approach, the analysis uncovered that there is preliminary evidence of an internal grammatical structure of complex constructions containing multiple intensifiers. This is a great candidate for future research with significant impact as well on future analyses of compounding. Moreover, some intensifiers occur relatively far more frequently in combination with other intensifiers than on their own with *geil*. For example, over 60% of all *geil* constructions containing *affen-* also contain additional intensifiers (most often *titten-*), compared to only 7% of *geil* constructions containing *hammer-* (the most frequent construction overall), suggesting that certain constructions are potentially becoming lexicalised.

Chapter 4 also demonstrated, using longitudinal analysis, that the three senses of *geil* in the *NottDeuYTSch* corpus have undergone different trends in the frequency of use over the time frame covered by the corpus (2008-2018). There has been a dramatic decrease in the usage of *geil* with the sense of 'cool', with the frequency dropping by around 85% from 193,000 instances per million words (ipm) in 2008 to 29,000 ipm in 2018. However, the frequencies of the sense of *geil* to mean 'desperate for something' increased from 0 recorded uses

in 2008 and 2009 to around 800 ipm in 2011, where it has since remained stable. The sense of *geil* to mean 'sexually aroused' experienced a peak in 2010 at 579 ipm but the frequency has since slowly declined. The sharp decline in the use of *geil* to mean 'cool' provides concrete evidence of the lexical change of this word in German youth language, which had previously only been suggested in existing studies. To reinforce this finding, the chapter also provided evidence of lexical items that are similarly used by young people to express a positive evaluation where usage has increased. The frequency of the word *cool* in the *NottDeuYTSch* corpus overtook the frequency of *geil* in 2014, and the corpus data also suggest that *mega* and *nice* may soon supersede *geil* as positive evaluations in German youth language. The considerable increase in the use of these words may offer fertile ground for future research on such expressions. These lexical and morphological findings have significant and multifaceted value. First, they demonstrate the value of quantitative and longitudinal approaches alongside qualitative study, even for a word as familiar in the research literature as *geil*. Second, they provide evidence for the existence of rapid lexical change in youth language, both in terms of the rise and fall of the use of *geil* and the patterns of new lexical items, such as *mega* and *nice*, that seem to be replacing it when used in the sense of 'cool'. This is further confirmation that youth language is constantly undergoing lexical innovation and renewal.

Chapters 5 and 6, investigated parataxis in clauses introduced by subordinating conjunctions. Studies to date had identified three subordinating conjunctions or conjunction clusters that are used to introduce paratactic clauses: *dass*; *obwohl*; and most well-known, *weil*, but Chapter 5 presented evidence of three additional subordinating conjunctions or conjunction clusters that are used to introduce paratactic clauses: *solange*; *wenn*; and *wenn* clusters (*auch wenn*, *selbst wenn*, *sogar wenn*, and *und wenn*), that have not been identified in previous research and suggests that the relationship between conjunctions and sentence structure in German syntax must be re-evaluated. Furthermore, previous research, outside of *weil*, has focused on *obwohl*, although the frequency of paratactic *obwohl* clauses compared to other subordinating conjunctions in the *NottDeuYTSch* corpus is relatively low. This suggests that the phenomenon

of parataxis and non-standard syntax has changed over the last decade and requires further research.

The longitudinal analysis of the data presented in Chapter 5 showed that the frequency of paratactic structures following subordinating clauses doubled from 2008 to 2018, the timespan of the *NottDeuYTSch* corpus. To date, research had merely implied that this syntactic phenomenon had been increasing, but this study has provided quantitative evidence confirming the trend. This increase was overwhelmingly fuelled by an increase in paratactic *weil* clauses, which comprise over 90% of all paratactic clauses introduced by a subordinating conjunction in the *NottDeuYTSch* corpus. By comparing the *NottDeuYTSch* corpus with three other corpora of written language, Chapter 5 also demonstrated that subordinating conjunctions introducing paratactic clauses occurred almost 400 times more frequently in digital youth language than in formal written German, and confirmed that higher use of this phenomenon correlates with the informality of the communicative environment and also potentially the age of the writer. This provides quantitative evidence to confirm previous theories, particularly the link between formality and non-standard linguistic features (e.g. Durrell 2003: 22). However, these findings also demonstrate the nuanced differences between sites of DMC, which is far less researched.

In Chapter 6, the closer analysis of *NottDeuYTSch* corpus data demonstrated that several hypotheses previously proposed to explain the use of paratactic structure following subordinating conjunctions are not – or no longer – applicable to youth language, e.g. the illocutionary autonomy or syntactic disintegration of the *weil* clause from the main clause, or the use of *weil* analogous to *denn*. Instead, analysis presented in Chapter 6 supports the hypothesis that several subordinating conjunctions have undergone grammaticalization, which allows them to function additionally as either a coordinating conjunction, a discourse marker, or both. This was particularly evident in clauses with predicative elements in the first field, e.g. *dann* and *eigentlich*, which occur with relatively far higher frequency in the first field of a paratactic *weil* clause than the subject. This was shown to be an indicator that the subordinating conjunction has a discourse structuring function, rather than introducing a causal link between two

clauses, and these findings with *weil* could also be applied to other subordinating conjunctions that have undergone similar grammaticalization. The chapter also demonstrated that several subordinating conjunctions, such as *trotzdem* and *als ob*, are used as discourse markers but not as coordinating conjunctions in the *NottDeuYTSch* corpus. This finding contradicts the linear model of grammaticalization presented by Gohl & Günthner (1999) and Günthner (1999), which posited that a subordinating conjunction first undergoes grammaticalization, allowing it to function as a coordinating conjunction, before undergoing a second grammaticalization process, allowing it to function as a discourse marker. Instead, Chapter 6 presents an alternative model of grammaticalization, according to which a subordinating conjunction may undergo grammaticalization, allowing it to function *either* as a coordinating conjunction *or* directly as a discourse marker, before undergoing a possible second grammaticalization process. The findings presented in Chapters 5 and Chapter 6 thus demonstrate that mixed methods corpus-linguistic approaches to syntax can uncover emergent developments and make it possible to re-examine and revise existing hypotheses using new data. Furthermore, grammaticalization was not shown to be linear, which impacts how we view the relationship between subordinating conjunctions, discourse markers and coordinating conjunctions.

Chapter 7 examined the orthographic strategies used to alter metacommunicative functions of digital messages and used the *NottDeuYTSch* corpus data to develop a framework for the analysis of metacommunication in Digital Writing. The chapter identified twelve groups of orthographic features with metacommunicative function, and, in particular, documented new features that have arisen as a result of technological developments and have replaced existing features, such as emoji (among other features) replacing emoticons. Within the groups of features, the chapter identified emergent orthographic strategies in the *NottDeuYTSch* corpus data, such as the adoption of features specific to other online platforms and their subsequent re-conventionalisation, e.g. the use of *kappa* to signal irony, that originated from the streaming website *Twitch* as an 'emote'.

Using the features identified earlier in the Chapter, Section 7.4 presented a

new unified, comprehensive and flexible framework to describe and analyse orthographic strategies used in Digital Writing for metacommunication, drawing on studies of DMC, pragmatics, and visual culture. The framework encompasses four overarching metacommunicative functions that could be transmitted through the use of orthographic features: illocutionary force, indexing identity, illustration, and structuring the information within the digital space; and captures the fact that orthographic features may simultaneously convey multiple metacommunicative functions. The importance of metacommunication in Digital Writing cannot be understated, as demonstrated in the wide-reaching examples in Chapter 7, yet a holistic framework to analyse it has not been found in previous research. Instead, previous frameworks have chiefly focused either on more general functions of orthographic features in Digital Writing or on the functions (although not necessarily metacommunicative) of more specific groups of orthographic features. The framework is therefore a significant contribution to the field of DMC, as it enables analyses of metacommunication in all kinds of Digital Writing. In particular, the development of the framework draws attention to how manipulating the space in which a digital text is written can affect the metacommunicative functions of the text. This aspect of analysis has not been previously addressed in existing studies of DMC, and is therefore a significant addition to the arsenal of analytical methods available to future researchers.

The framework further demonstrates that DMC is not just an imitation of spoken language, as per the popular theory in the 1990s and 2000s, as it provides examples of metacommunicative functions that are 'born digital', such as the features listed under the literal representation of visual metacommunication in Section 7.1.3. Additionally, the framework enables nuanced interpretations of the functions of orthographic features in Digital Writing, particularly features that have multiple potential interpretations or micro-variation within features that can change the function, e.g. the use of different skin tones with certain emoji can have different metacommunicative functions. While the framework was developed using the text-based data of the *NottDeuYTSch* corpus, it is also applicable to other areas of DMC, such as multimodal digital interaction, which is rapidly

increasing in prevalence. Furthermore, it may be a useful framework for more in-depth analyses of certain orthographic features, such as emoji ethnography, which, I anticipate, will become a productive area of study in the near future.

In sum, the analyses presented in this thesis of data from the 33-million word *NottDeuYTSch* corpus have demonstrated that digital youth language is a rich source of lexical, morphological, syntactic, and orthographical innovation. The mixed methods corpus-linguistic approaches, variously combining quantitative, longitudinal, and qualitative methodologies have uncovered significant developments and emergent features in the language of young German speakers in *YouTube* comments, addressing several gaps in previous scholarship on linguistic developments in both standard and youth varieties of German, and – especially in Chapter 7 – on Digital Writing more widely, irrespective of age or language. The three case studies presented here – addressing questions of lexis and morphology; syntax; and digital writing – showcase the breadth of linguistic analysis that can be achieved using the *NottDeuYTSch* corpus, and demonstrate the significant and extensive possibilities for corpus-linguistic analysis of the rapidly changing and creative language variety that is German youth language.

# Part IV

# Bibliography and Appendix

# Bibliography

1Live. 2019a. 'Babo-Bus: Gameshows | 1live - YouTube' <https://www.youtube.com/watch?v=kBT4PxaHzRI> [accessed 21 February 2019]

———. 2019b. 'Der 1LIVE Babo-Bus: Gameshows,' *WDR Nachrichten* <https://www1.wdr.de/radio/1live/comedy/babo-bus/video-der-live-babo-bus-gameshows-100.html> [accessed 21 February 2019]

Abbott, Chris. 1998. 'Making Connections: Young People and the Internet,' in *Digital Diversions: Youth Culture in the Age of Multimedia*, ed. by Julian Sefton-Green (London: Routledge), pp. 79–98

Abidin, Crystal. 2019. 'Yes Homo: Gay Influencers, Homonormativity, and Queerbaiting on YouTube,' *Continuum*, 33.5 (Taylor & Francis): 614–29

Ahn, Wonmi, Jeeyea Park, & Kwang-hee Han. 2011. 'Emoticons Convey Emotion in CMC,' in *Proceedings of the 25th BCS Conference on Human-Computer Interaction* (Swindon: British Computer Society), pp. 429–30

Albert, Georg. 2015. 'Semiotik Und Syntax von Emoticons,' *Zeitschrift für Angewandte Linguistik*, 62.1: 3–22

Alexa Internet, Inc. 2021. 'Alexa Top 500 Global Sites' <http://www.alexa.com/topsites> [accessed 27 September 2019]

Ammon, Ulrich. 2005. 'Standard Und Variation: Norm, Autorität, Legitimation,' in *Standardvariation: Wie Viel Variation Verträgt Die Deutsche Sprache?*, ed.

by Ludwig Eichinger & Werner Kallmeyer (Berlin: Walter de Gruyter), pp. 28–40

Anderson, Benedict. 1983. *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (London: Verso)

Anderson, Jeffrey F, Fred K Beard, & Joseph B Walther. 2010. 'Turn-Taking and the Local Management of Conversation in a Highly Simultaneous Computer-Mediated Communication System,' *Language@Internet*, 7

Androutsopoulos, Jannis. 1998. *Deutsche Jugendsprache: Untersuchungen Zu Ihren Strukturen Und Funktionen* (Frankfurt am Main: Peter Lang)

———. 2000. 'Non-Standard Spellings in Media Texts: The Case of German Fanzines,' *Journal of Sociolinguistics*, 4: 514–33 <https://doi.org/10.1111/1467-9481.00128>

———. 2001. 'Ultra Korregd Alder! Zur Medialen Stilisierung Und Popularisierung von "Türkendeutsch",' *Deutsche Sprache*, 29: 321–39

——— (ed.). 2003a. *HipHop: Globale Kultur-Lokale Praktiken* (Bielefeld: Transcript)

———. 2003b. 'Online-Gemeinschaften Und Sprachvariation. Soziolinguistische Perspektiven Auf Sprache Im Internet,' *Zeitschrift für Germanistische Linguistik*, 31: 173–97

———. 2005. '*... 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 & 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 <https://doi.org/10.1111/j.1467-9841.2006.00291.x>

———. 2007. 'Neue Medien–Neue Schriftlichkeit,' *Mitteilungen Des Deutschen Germanistenverbandes*, 1.7: 72–97

———. 2008. 'Research on Youth-Language / Jugendsprach-Forschung,' in

*Sociolinguistics/Soziolinguistik*, HSK 3.2, 2nd edn, ed. by Ulrich Ammon et al. (Berlin: Walter de Gruyter), pp. 1496–1505 <https://doi.org/10.1515/9783110171488.2>

———. 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

———. 2011. 'Language Change and Digital Media: A Review of Conceptions and Evidence,' in *Standard Languages and Language Standards in a Changing Europe*, Standard Language Ideology in Contemporary Europe, ed. by Tore Kristiansen & Nikolas Coupland (Oslo: Novus Press), pp. 145–61

———. 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 & 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*, 19.2: 185–205 <https://doi.org/10.1177/1367006913489198>

———. 2018. 'Digitale Interpunktion: Stilistische Ressourcen und soziolinguistischer Wandel in der informellen digitalen Schriftlichkeit von Jugendlichen,' in *Jugendsprachen: Aktuelle Perspektiven Internationaler Forschung*, ed. by Arne Ziegler (Berlin: De Gruyter), pp. 721–48

———. 2020a. 'Auslassungspunkte in Der Schriftbasierten Interaktion,' in *Register Des Graphischen*, ed. by Jannis Androutsopoulos & Florian Busch (De Gruyter), pp. 133–58 <https://doi.org/10.1515/9783110673241-006>

———. 2020b. 'Digitalisierung und soziolinguistischer Wandel: Der Fall der digitalen Interpunktion,' in *Deutsch in Sozialen Medien: Interaktiv – multimodal – vielfältig*, ed. by Konstanze Marx, Henning Lobin, & Axel Schmidt (De Gruyter), pp. 75–94 <https://doi.org/10.1515/9783110679885-005>

Androutsopoulos, Jannis, & Florian Busch (eds.). 2020. *Register des Graphischen: Variation, Interaktion und Reflexion in der digitalen Schriftlichkeit* (De Gruyter)

———. 2021. 'Digital Punctuation as an Interactional Resource: The Message-Final Period Among German Adolescents,' *Linguistics and Education*, Digital language practices: media, awareness, pedagogy, 62: 100871 <https://doi.org/10.1016/j.linged.2020.100871>

Androutsopoulos, Jannis, & Alexandra Georgakopoulou. 2003. 'Discourse Constructions of Youth Identities: Introduction,' in *Discourse Constructions of Youth Identities*, ed. by Jannis Androutsopoulos & Alexandra Georgakopoulou (Philadelphia: John Benjamins), cx, pp. 1–26

Androutsopoulos, Jannis, & Gurly Schmidt. 2002. 'SMS-Kommunikation: Ethnografische Gattungsanalyse am Beispiel einer Kleingruppe,' *Zeitschrift für Angewandte Linguistik*: 49–80

Androutsopoulos, Jannis, & Arno Scholz (eds.). 1998. *Jugendsprache, Langue Des Jeunes, Youth Language* (Frankfurt am Main: Peter Lang)

Androutsopoulos, Jannis, & Jana Tereick. 2016. 'YouTube: Language and Discourse Practices in Participatory Culture,' in *The Routledge Handbook of Language and Digital Communication*, Routledge Handbooks in Applied Linguistics, ed. by Alexandra Georgakopoulou & Tereza Spilioti (Abingdon: Routledge), pp. 354–70

Antomo, Mailin, & Markus Steinbach. 2010. 'Desintegration Und Interpretation: Weil-V2-Sätze an Der Schnittstelle Zwischen Syntax, Semantik Und Pragmatik,' *Zeitschrift für Sprachwissenschaft*, 29.1: 1–37 <https://doi.org/10.1515/zfsw.2010.001>

Arens, Katja. 2019. 'Bild-Makros in Der Facebook-Interaktion. Eine Medienlinguistische Betrachtung Multimodaler Kommunikate Und Ihrer Interaktiven Aushandlung,' *NetWorx* <https://doi.org/10.15488/4654>

Aschwanden, Brigitte. 2001. '»Wär Wot Chätä?« Zum Sprachverhalten Deutschschweizerischer Chatter,' *NetWorx*

<https://doi.org/10.15488/2915>

Ascoop, Kristin, & Torsten Leuschner. 2005. 'Affixoidhungrig? Skitbra!' *Germanistische Mitteilungen*, 62: 1–13

Aslan, Erhan, & Camilla Vásquez. 2018. '"Cash Me Ousside": A Citizen Sociolinguistic Analysis of Online Metalinguistic Commentary,' *Journal of Sociolinguistics*, 22.4: 406–31

Auer, Peter. 1997. 'Formen Und Funktionen Der Vor-Vorfeldbesetzung Im Gesprochenen Deutsch,' in *Syntax Des Gesprochenen Deutsch*, ed. by Peter Schlobinski (Wiesbaden: VS Verlag für Sozialwissenschaften), pp. 55–91 <https://doi.org/10.1007/978-3-322-88924-9_3>

———. 1998. 'Zwischen Parataxe und Hypotaxe: "abhängige Hauptsätze" im Gesprochenen und Geschriebenen Deutsch,' *Zeitschrift für Germanistische Linguistik*, 26.3: 284–307

———. 2003. '"Türkenslang": Ein Jugendsprachlicher Ethnolekt Des Deutschen Und Seine Transformationen,' in *Spracherwerb Und Lebensalter*, ed. by Annelies Häcki Buhofer (Tübingen: Franke), pp. 255–64

———. 2016. '„Wie geil ist das denn?",' *Zeitschrift für germanistische Linguistik*, 44.1 (De Gruyter): 69–92 <https://doi.org/10.1515/zgl-2016-0003>

Auer, Peter, & 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 & Alexandra Georgakopoulou (Philadelphia: John Benjamins), pp. 223–46

Auer, Peter, & Susanne Günthner. 2003. 'Die Entstehung von Diskursmarkern Im Deutschen – Ein Fall von Grammatikalisierung,' in *Grammatikalisierung Im Deutschen*, ed. by Torsten Leuschner, Tanja Mortelmans, & Sarah De Groodt (Berlin: De Gruyter), pp. 335–62

Aull, Bethany. 2019. 'A Study of Phatic Emoji Use in WhatsApp Communication,' *Internet Pragmatics*, 2.2 (John Benjamins): 206–32

<https://doi.org/10.1075/ip.00029.aul>

Austin, John Langshaw. 1962. *How to Do Things with Words* (London: Oxford University Press)

Autenrieth, Ulla, & Klaus Neumann-Braun. 2016. 'Immer Vernetzt - Peerbeziehungen von Jugendlichen in Online-Umgebungen,' in *Handbuch Peerforschung*, ed. by Sina-Mareen Köhler, Heinz-Hermann Krüger, & Nicolle Pfaff (Leverkusen: Babara Budrich), pp. 339–52

Auxland, Morrigan. 2020. 'Para Todes: A Case Study on Portuguese and Gender-Neutrality,' *Journal of Languages, Texts, and Society*: 60–83

Ágel, Vilmos. 2016. '*Obwohl (.) Fährt Der Eigentlich Auch Am Sonntag?* Der Verbzweit-Mythos,' in *Gesprochene Fremdsprache Deutsch*, ed. by Brigitte Handwerker, Reiner Bäuerle, & Bernd Sieberg (Baltmannsweiler: Schneider Verlag), pp. 75–100

Ágel, Vilmos, & Mathilde Hennig. 2006. 'Überlegungen zur Theorie und Praxis des Nähe- und Distanzsprechens,' in *Überlegungen zur Theorie und Praxis des Nähe- und Distanzsprechens*, ed. by Vilmos Ágel & Mathilde Hennig (Tübingen: Max Niemeyer Verlag), pp. 179–214 <https://www.degruyter.com/document/doi/10.1515/9783110936063.179/html> [accessed 2 June 2021]

Bader, Jennifer. 2002. 'Schriftlichkeit Und Mündlichkeit in Der Chat-Kommunikation,' *NetWorx*, 29

Bahlo, Nils. 2012. 'Let's Talk about Sex–Vulgärer Und Sexualisierter Sprachgebrauch Jugendlicher Als Thema Im Projektunterricht,' *Aptum. Zeitschrift für Sprachkritik Und Sprachkultur*, 8.1: 48–60

Bahlo, Nils et al. 2019. *Jugendsprache: Eine Einführung* (Berlin: J.B. Metzler)

Bahlo, Nils. 2021. 'Zur multimodalen Indizierung von Postadoleszenz - Gedanken und Beispiele zur konstruktionsgrammatischen Beschreibung ausgewählter Altersmarker in der Interaktion' (presented at the Kolloquium Mehrsprachigkeit, Sprachkontakt, Variation, Berlin: Humboldt University)

Bahlo, Nils, & Wolfgang Klein. 2017. '„Jugendsprache “,' in *Vielfalt Und Einheit Der Deutschen Sprache. Zweiter Bericht Zur Lage Der Deutschen Sprache.*, ed. by Deutsche Akademie für Sprache und Dichtung & Union der deutschen Akademien der Wissenschaften (Tübingen: Stauffenburg), pp. 145–90

Baker, Paul. 2010. *Sociolinguistics and Corpus Linguistics* (Edinburgh: Edinburgh University Press) [accessed 23 July 2018]

———. 2018. 'Corpus Methods in Linguistics,' in *Research Methods in Linguistics*, 2nd edn, ed. by Lia Litosseliti (London: Continuum), pp. 93–113

Balirano, Giuseppe. 2020. 'Of Rainbow Unicorns. The Role of Bonding Queer Icons in Contemporary LGBTIQ+ Re-Positionings,' *Ocula*, 21 <https://doi.org/10.12977/ocula2020-7>

Barbaresi, Adrien. 2019. 'The Vast and the Focused: On the Need for Thematic Web and Blog Corpora,' in *Proceedings of the Workshop on Challenges in the Management of Large Corpora*, ed. by Piotr Bański et al. (Mannheim: Leibniz-Institut für Deutsche Sprache), pp. 29–32

Barbaresi, Adrien, & Alexander Geyken. 2020. 'Die Webkorpora Im DWDS–Strategien Des Korpusaufbaus Und Nutzungsmöglichkeiten,' in *Deutsch in Sozialen Medien: Interaktiv – Multimodal – Vielfältig*, ed. by Konstanze Marx, Henning Lobin, & Axel Schmidt (Berlin: De Gruyter), pp. 345–48

Barbieri, Francesco et al. 2017. 'Towards the Understanding of Gaming Audiences by Modeling Twitch Emotes,' in *Proceedings of the 3rd Workshop on Noisy User-Generated Text* (presented at the Proceedings of the 3rd Workshop on Noisy User-generated Text, Copenhagen: Association for Computational Linguistics), pp. 11–20 <https://doi.org/10.18653/v1/W17-4402>

Barbieri, Francesco et al. 2018. 'Multimodal Emoji Prediction,' in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short*

*Papers)* (presented at the NAACL-HLT 2018, New Orleans: Association for Computational Linguistics), pp. 679–86 <https://doi.org/10.18653/v1/N18-2107>

Barbieri, Francesco, & Jose Camacho-Collados. 2018. 'How Gender and Skin Tone Modifiers Affect Emoji Semantics in Twitter,' in *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics* (presented at the Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics, New Orleans: Association for Computational Linguistics), pp. 101–6 <https://doi.org/10.18653/v1/S18-2011>

Barbieri, Francesco, Horacio Saggion, & Francesco Ronzano. 2014. 'Modelling Sarcasm in Twitter, a Novel Approach,' in *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (presented at the Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Baltimore: Association for Computational Linguistics), pp. 50–58 <https://doi.org/10.3115/v1/W14-2609>

Baron, Naomi S. 1984. 'Computer Mediated Communication as a Force in Language Change,' *Visible Language*, 18.2: 118–41 [accessed 8 February 2021]

———. 2004. 'See You Online: Gender Issues in College Student Use of Instant Messaging,' *Journal of Language and Social Psychology*, 23: 397–423

———. 2008. *Always on: Language in an Online and Mobile World* (Oxford: Oxford University Press)

Barton, David, & Carmen Lee. 2013. *Language Online: Investigating Digital Texts and Practices* (London: Routledge)

Bateson, Gregory. 1972. *Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology.* (Chicago: University of Chicago Press)

Bayer, Josef. 1984. 'Comp in Bavarian Syntax,' *Linguistic Review*, 3.3: 209–74

<https://doi.org/10.1515/tlir.1984.3.3.209>

Bazin, Hervé. 1966. *Plumons l'oiseau* (Paris: Editions Bernard Grasset)

Beck, Klaus. 2010. 'Soziologie Der Online-Kommunikation,' in *Handbuch Online-Kommunikation*, ed. by Wolfgang Schweiger & Klaus Beck (Wiesbaden: VS Verlag für Sozialwissenschaften), pp. 15–35

Beeman, William O. 2019. 'Paralinguistic and Performative Dimensions,' in *The Routledge Handbook of Language and Emotion* (Abingdon: Routledge)

Beißwenger, Michael et al. 2015. 'Adding Value to CMC Corpora: CLARINification and Part-of-Speech Annotation of the Dortmund Chat Corpus,' in *Proceedings of the 2nd Workshop on Natural Language Processing for Computer-Mediated Communication / Social Media*, pp. 12–16

Beißwenger, Michael et al. 2020. 'Die Mobile Communication Database 2 (MoCoDa 2),' in *Deutsch in Sozialen Medien: Interaktiv – Multimodal – Vielfältig*, ed. by Konstanze Marx, Henning Lobin, & Axel Schmidt (Berlin: De Gruyter), pp. 349–52

Beißwenger, Michael, & Steffen Pappert. 2019a. *Handeln Mit Emojis: Grundriss Einer Linguistik Kleiner Bildzeichen in Der WhatsApp-Kommunikation* (Duisburg: Universitätsverlag Rhein-Ruhr)

———. 2019b. 'How to Be Polite with Emojis: A Pragmatic Analysis of Face Work Strategies in an Online Learning Environment,' *European Journal for Applied Linguistics*, 7.2: 225–53 <https://www.degruyter.com/view/journals/eujal/7/2/article-p225.xml> [accessed 19 November 2020]

———. 2020. 'Small Talk Mit Bildzeichen,' *Zeitschrift für Literaturwissenschaft Und Linguistik*, 50.1: 89–114 <https://doi.org/10.1007/s41244-020-00160-5>

Beißwenger, Michael, & Angelika Storrer. 2012. 'Interaktionsorientiertes Schreiben und interaktive Lesespiele in der Chat-Kommunikation,'

*Zeitschrift für Literaturwissenschaft und Linguistik*, 42.4: 92–124 <https://doi.org/10.1007/BF03391261>

Bell, Diana C., & Mike T. Hübler. 2001. 'The Virtual Writing Center: Developing Ethos Through Mailing List Discourse,' *The Writing Center Journal*, 21.2: 57–78

Ben Moussa, Mohamed. 2019. 'Rap It up, Share It up: Identity Politics of Youth "Social" Movement in Moroccan Online Rap Music,' *New Media & Society*, 21.5: 1043–64 <https://doi.org/10.1177/1461444818821356>

Beneke, Jürgen. 1985. 'Zur Sozialen Differenziertheit Der Sprache Am Beispiel Jugendtypischer Sprechweise,' *STUF-Language Typology and Universals*, 38.1-6 (De Gruyter Akademie Forschung): 251–63

Benenson, Fred. 2010. *Emoji Dick; or the Whale* (Morrisville: Lulu Press)

Berend, Nina. 2005. 'Regionale Gebrauchsstandards–Gibt Es Sie Und Wie Kann Man Sie Beschreiben,' in *Standardvariation. Wie Viel Variation Verträgt Die Deutsche Sprache* (Berlin: De Gruyter), pp. 143–70

Berg, Kristian. 2020. 'Variation Im Digitalen Schreiben,' in *Register Des Graphischen*, ed. by Jannis Androutsopoulos & Florian Busch (De Gruyter), pp. 53–66 <https://doi.org/10.1515/9783110673241-003>

Biber, Douglas. 1988. *Variation Across Speech and Writing* (Cambridge: Cambridge University Press)

———. 1993. 'Representativeness in Corpus Design,' *Literary and Linguistic Computing*, 8: 243–57

Bierbach, Christine, & 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

Bildhauer, Felix, & Philippa Cook. 2010. 'German Multiple Fronting and Expected Topic-Hood,' in *Proceedings of the 17th International Conference on Head-Driven Phrase Structure Grammar, Université Paris Diderot*

(Stanford: CSLI Publications), pp. 68–79

Blashki, Katherine, & Sophie Nichol. 2005. 'Game Geek's Goss: Linguistic Creativity in Young Males Within an Online University Forum (94/\/\3 933k'5 9055oneone),' *Australian Journal of Emerging Technologies and Society*, 3.2 (Australian Centre for Emerging Technologies and Society): 71–80

Blommaert, Jan, Elina Westinen, & Sirpa Leppänen. 2015. 'Further Notes on Sociolinguistic Scales,' *Intercultural Pragmatics*, 12.1 <https://doi.org/10.1515/ip-2015-0005>

Blue, Violet. 2013. 'Forced Google Plus Integration on YouTube Backfires, Petition Hits 112,000,' *ZDNet* <https://www.zdnet.com/article/forced-google-plus-integration-on-youtube-backfires-petition-hits-112000/> [accessed 5 March 2021]

Blühdorn, Hardarik, Ad Foolen, & Oscar Loureda. 2017. 'Diskursmarker: Begriffsgeschichte – Theorie – Beschreibung: Ein Bibliographischer Überblick,' in *Diskursmarker Im Deutschen. Reflexionen Und Analysen.*, ed. by Hardarik Blühdorn et al. (Göttingen: Verlag für Gesprächsforschung), pp. 7–47 <https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/6287/file/Bluehdorn_Foolen_Loreda_Diskursmarker_2017.pdf> [accessed 29 May 2020]

Bohn, Willard. 2001. *Modern Visual Poetry* (London: Associated University Presses)

———. 2010. *Reading Visual Poetry* (Plymouth: Fairleigh Dickinson University Press)

Bonnytrash. 2018. 'Bonnytrash Official Shop' <http://www.bonny-shop.de> [accessed 23 April 2021]

Borba, Rodrigo. 2019. 'The Interactional Making of a "True Transsexual": Language and (Dis)identification in Trans-Specific Healthcare,' *International Journal of the Sociology of Language.*, 2019.256: 21–55

Bou-Franch, Patricia, & Pilar Garcés-Conejos Blitvich. 2014. 'Gender Ideology and Social Identity Processes in Online Language Aggression Against Women,' *Journal of Language Aggression and Conflict*, 2: 226–48

Bou-Franch, Patricia, Nuria Lorenzo-Dus, & Pilar Garcés-Conejos Blitvich. 2012. 'Social Interaction in YouTube Text-Based Polylogues: A Study of Coherence,' *Journal of Computer-Mediated Communication*, 17.4: 501–21

Böckmann, Barbara et al. 2019. 'Multimodale Produktbewertungen in Videos von Influencerinnen Auf YouTube: Zur Parainteraktiven Konstruktion von Warenwelten,' *Zeitschrift für Angewandte Linguistik*, 70.1 (De Gruyter Mouton): 139–72

Breindl, Eva. 2008. 'Die Brigitte Nun Kann Der Hans Nicht Ausstehen: Gebundene Topiks Im Deutschen,' *Deutsche Sprache*: 27–49 <https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/6327/file/Breindl_Die_Brigitte_nun_kann_der_Hans_nicht_ausstehn_2008.pdf> [accessed 21 August 2020]

———. 2012. 'Er Ist Sympathisch, Weil Menschlich – Weil Als Koordinierende Konjunktion?' in *Grammatische Stolpersteine Digital: Festschrift für Bruno Strecker Zum 65. Geburtstag*, ed. by Marek Konopka, Roman Schneider, & Bruno Strecker (Mannheim: Leibniz-Institut für Deutsche Sprache), pp. 153–58

Breindl, Eva, Anna Volodina, & Ulrich Hermann Waßner. 2014. *Handbuch der deutschen Konnektoren 2, Semantik der deutschen Satzverknüpfer* (Berlin: De Gruyter) <https://doi.org/10.1515/9783110341447>

Brommer, Sarah. 2007. '"Ein Unglaubliches Schriftbild, von Rechtschreibung Oder Interpunktion Ganz Zu Schweigen": Die Schreibkompetenz Der Jugendlichen Im öffentlichen Diskurs,' *Zeitschrift für Germanistische Linguistik*, 35.3 (De Gruyter): 315–45

Brommer, Sarah, & Christa Dürscheid. 2012. 'Mediennutzung heutiger Jugendlicher: Generation Facebook?' in *Sprache der Generationen*, ed. by Eva Neuland (Mannheim: Dudenverlag), pp. 271–93

<https://www.zora.uzh.ch/id/eprint/54270> [accessed 27 July 2021]

Bross, Fabian. 2012. 'German Modal Particles and the Common Ground,' *Helikon. A Multidisciplinary Online Journal*, 2: 182–209

Bruce & Bongo. 1986. *'Geil' (Geilomatick Mix)* (Munich: Ariola)

Bunk, Oliver. 2020. '"Aber immer alle sagen das" The Status of V3 in German: Use, Processing, and Syntactic Representation' (Berlin: Humboldt-Universität zu Berlin) <https://edoc.hu-berlin.de/handle/18452/22779> [accessed 16 October 2021]

Burge, Jeremy. 2019a. 'Correcting the Record on the First Emoji Set,' *Emojipedia* <https://blog.emojipedia.org/correcting-the-record-on-the-first-emoji-set/> [accessed 14 April 2021]

———. 2019b. 'Priorities for Future RGI Emoji Sequences' (Unicode Consortium) <https://www.unicode.org/L2/L2019/19101-esc-rgi-priorities.pdf>

Burgess, Jean, & Joshua Green. 2018. *YouTube: Online Video and Participatory Culture*, 2nd edn (Cambridge: Polity Press)

Burkhardt, Armin. 1990. 'Speech Act Theory - the Decline of a Paradigm,' in *Speech Acts, Meaning and Intentions: Critical Approaches to the Philosophy of John R. Searle*, ed. by Armin Burkhardt (Berlin: Walter de Gruyter), pp. 91–126

———. 2013. 'Die "Anglizismen-Frage" Aus Der Sicht Der GfdS,' *Sprachreport. Informationen Und Meinungen Zur Deutschen Sprache*, 29.1/2: 38–42

Busch, Florian. 2017. 'Informelle Interpunktion? Zeichensetzung Im Digitalen Schreiben von Jugendlichen,' *Der Deutschunterricht*, 4: 87–91

———. 2018. 'Digitale Schreibregister von Jugendlichen analysieren - Ein linguistisch-ethnographischer Zugang zu Praktiken des Alltagsschreibens,' in *Jugendsprachen: Aktuelle Perspektiven Internationaler Forschung*, ed. by Arne Ziegler (Berlin: De Gruyter), pp. 829–58

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

Cameron, Deborah. 1992. *Feminism and Linguistic Theory*, 2nd edn (London: Palgrave Macmillan)

Canagarajah, Suresh. 2011. 'Codemeshing in Academic Writing: Identifying Teachable Strategies of Translanguaging,' *Modern Language Journal*, 95: 401–17

Carey, John. 1980. 'Paralanguage in Computer Mediated Communication,' in *Proceedings of the 18th Annual Meeting on Association for Computational Linguistics* - (presented at the the 18th annual meeting, Philadelphia: Association for Computational Linguistics), p. 67 <https://doi.org/10.3115/981436.981458>

Carr, Caleb T. 2020. 'CMC Is Dead, Long Live CMC!: Situating Computer-Mediated Communication Scholarship Beyond the Digital Age,' *Journal of Computer-Mediated Communication*, 25.1: 9–22 [accessed 4 February 2021]

Castor, Theresa. 2017. 'Metacommunication,' in *The International Encyclopedia of Organizational Communication*, ed. by Craig R. Scott et al. (Oxford: Wiley), pp. 1–9 <https://doi.org/10.1002/9781118955567>

Chau, Clement. 2010. 'YouTube as a Participatory Culture,' *New Directions for Youth Development*, 2010.128: 65–74 <https://doi.org/https://doi.org/10.1002/yd.376>

Chen, Zhenpeng et al. 2018. 'Through a Gender Lens: Learning Usage Patterns of Emojis from Large-Scale Android Users,' in *Proceedings of the 2018 World Wide Web Conference*, WWW '18 (Geneva: International World Wide Web Conferences Steering Committee), pp. 763–72 <https://doi.org/10.1145/3178876.3186157>

Chun, Elaine W. 2013. 'Ironic Blackness as Masculine Cool: Asian American Language and Authenticity on YouTube,' *Applied Linguistics*, 34: 592–612 <https://doi.org/10.1093/applin/amt023>

Chun, Elaine, & Keith Walters. 2011. 'Orienting to Arab Orientalisms: Language, Race, and Humor in a YouTube Video,' in *Digital Discourse: Language in the New Media*, ed. by Crispin Thurlow & Kristine R. Mroczek (Oxford: Oxford University Press), pp. 251–73

Clammer, John. 2015. 'Performing Ethnicity: Performance, Gender, Body and Belief in the Construction and Signalling of Identity,' *Ethnic and Racial Studies*, 38: 2159–66 <https://doi.org/10.1080/01419870.2015.1045305>

Colliver, Ben, & Adrian Coyle. 2020. '"Risk of Sexual Violence Against Women and Girls" in the Construction of "Gender-Neutral Toilets": A Discourse Analysis of Comments on YouTube Videos,' *Journal of Gender-Based Violence*, 4.3 (Policy Press): 359–76

Collot, Milena, & Nancy Belmore. 1996. 'Electronic Language: A New Variety of English,' in *Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural Perspectives*, ed. by Susan C Herring (Amsterdam: John Benjamins), pp. 13–28

Cotgrove, Louis Alexander. 2013. 'The Influx of Anglicisms in German Internet Youth Language: A Study of Web Fora on Www.bravo.de' (unpublished Bachelor's Thesis, Nottingham: University of Nottingham)

———. 2017. '#GlockeAktiv: Gender and Ethnicity Differences in German-Language YouTube Comments' (unpublished Master's Thesis, Nottingham: University of Nottingham)

———. 2018. 'The Importance of Linguistic Markers of Identity and Authenticity in German Gangsta Rap,' *Journal of Languages, Texts, and Society*, 2: 67–98

Cotgrove, Louis Alexander, Kathryn Conklin, & Rüdiger Thul, in prep. 'The Conversationality Index: A Computational Assessment of Conversation in Social Media Interactions,' *Discourse Processes*

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. 2004. *A Glossary of Netspeak and Textspeak* (Edinburgh: Edinburgh University Press)

———. 2006. *Language and the Internet*, 2nd edn (Cambridge: Cambridge University Press)

———. 2008. 'Texting,' *ELT Journal*, 62.1: 77–83

———. 2011. *Internet Linguistics: A Student Guide* (Abingdon: Routledge)

Cueva, Lovón et al. 2021. 'Lenguaje, racismo y poder en el YouTube: Representaciones hegemónicas sobre los parlamentarios cultos peruanos,' *Discurso & Sociedad*: 348–82 <http://www.dissoc.org/ediciones/v15n02/DS15(2)Lovon&Quispe.pdf> [accessed 28 July 2021]

Cunningham, Stuart, & David Craig. 2017. 'Being 'Really Real'on YouTube: Authenticity, Community and Brand Culture in Social Media Entertainment,' *Media International Australia*, 164.1: 71–81

Curtis, Pavel. 1992. 'Mudding: Social Phenomena in Text-Based Virtual Realities,' *High Noon on the Electronic Frontier: Conceptual Issues in Cyberspace*: 347–74

Cutler, Cecelia. 2016. '"Ets Jast Ma Boooooooooooo": Social Meanings of Scottish Accents on YouTube,' in *English in Computer-Mediated Communication* (De Gruyter Mouton), pp. 69–98

———. 2018. '"Pink Chess Gring Gous": Discursive and Orthographic Resistance Among Mexican-American Rap Fans on YouTube,' in *Multilingual Youth Practices in Computer Mediated Communication*, ed. by Cecelia Cutler & Unn Røyneland (Cambridge: Cambridge University Press), pp. 127–44 <https://doi.org/10.1017/9781316135570.008>

———. 2020. 'Metapragmatic Comments and Orthographic Performances of a New York Accent on YouTube,' *World Englishes*, 39.1 (Wiley Online Library): 36–53

Çelikten, Esra, & Hakan Çelikten. 2020. 'Elektronik kültür Ortaminda

Neolojizm,' *Uluslararası Dil Edebiyat Ve Kültür Araştırmaları Dergisi*, 3.2: 183–205 <https://doi.org/10.37999/udekad.785222>

Daer, Alice R., Rebecca F. Hoffman, & Seth Goodman. 2015. 'Rhetorical Functions of Hashtag Forms Across Social Media Applications,' *Commun. Des. Q. Rev*, 3.1: 12–16 <https://doi.org/10.1145/2721882.2721884>

Dainas, Ashley R, & Susan C Herring. 2020. 'Interpreting Emoji Pragmatics,' in *Approaches to Internet Pragmatics: Theory and Practice*, ed. by Chaoqun Xie, Francisco Yus, & Hartmut Haberland (Amsterdam: John Benjamins) <http://ella.ils.indiana.edu/~herring/Interpreting_Emoji_Pragmatics.pdf>

Dander, Valentin. 2020. 'Generation Youtube: Die Zerstörung der Politik oder die Fortsetzung von Politik mit anderen Mitteln,' in *Jugendliche Mediennutzung und die Zukunft des Qualitätsjournalismus*, 1st edn, ed. by Katharina Gallner-Holzmann, Theo Hug, & Günther Pallaver (Innsbruck: Innsbruck university press), pp. 85–98 <https://doi.org/10.15203/99106-014-7-10>

Danesi, Marcel. 2016. *The Semiotics of Emoji: The Rise of Visual Language in the Age of the Internet* (Bloomsbury Publishing)

———. 2019. 'Emojis: Langue or Parole?' *Chinese Semiotic Studies*, 15.2 (De Gruyter Mouton): 243–58 <https://doi.org/10.1515/css-2019-0015>

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

Davidov, Dmitry, Oren Tsur, & Ari Rappoport. 2010. 'Semi-Supervised Recognition of Sarcastic Sentences in Twitter and Amazon,' in *Proceedings of the Fourteenth Conference on Computational Natural Language Learning* (Uppsala: Association for Computational Linguistics), pp. 107–16

De Decker, Benny, & Reinhild Vandekerckhove. 2017. 'Global Features of Online Communication in Local Flemish: Social and Medium-Related Determinants,' *Folia Linguistica*, 51.1 (De Gruyter): 253–81

<https://doi.org/10.1515/flin-2017-0007>

Deppermann, Arnulf. 2007. 'Stilisiertes Türkendeutsch in Gesprächen Deutscher Jugendlicher,' *Zeitschrift für Literaturwissenschaft Und Linguistik*, 37.4: 43–62

Der Tourist. 2013. *Supergeil (Feat. Friedrich Liechtenstein)* ([Merlin] FUGA Aggregation)

Diekmannshenke, Hajo. 2000. 'Die Spur Des Internetflaneurs: Elektronische Gästebücher Als Neue Kommunikationsform,' in *Soziales Im Netz: Sprache, Beziehungen Und Kommunikationskulturen Im Internet* (Wiesbaden: Westdeutscher Verlag), pp. 131–55

———. 2007. '*Lol* Gutes Deutsch in Neuen Medien,' in *Was Ist Gutes Deutsch?: Studien Und Meinungen Zum Gepflegten Sprachgebrauch*, ed. by Armin Burkhardt (Mannheim: Dudenverlag), pp. 213–27

Dieterle, Edward. 2009. 'Multi-User Virtual Environments for Teaching and Learning,' in *Encyclopedia of Multimedia Technology and Networking*, 2nd edn (IGI Global), pp. 1033–41

Dmitrow-Devold, Karolina. 2017. 'Performing the Self in the Mainstream: Norwegian Girls in Blogging,' *Nordicom Review*, 38.2: 65–78

Donalies, Elke. 2005. *Die Wortbildung Des Deutschen: Ein Überblick*, 2nd edn (Tübingen: Narr)

Dore, John. 1975. 'Holophrases, Speech Acts and Language Universals,' *Journal of Child Language*, 2.1: 21–40

Dowell, Nia M.M. et al. 2017. 'The Changing Patterns of MOOC Discourse,' in *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale* (presented at the L@S 2017: Fourth (2017) ACM Conference on Learning @ Scale, Cambridge, MA: ACM), pp. 283–86 <https://doi.org/10.1145/3051457.3054005>

Döring, Nicola. 2010. 'Sozialkontakte Online: Identitäten, Beziehungen, Gemeinschaften,' in *Handbuch Online-Kommunikation*, ed. by Wolfgang

Schweiger & Klaus Beck (Wiesbaden: VS Verlag für Sozialwissenschaften), pp. 159–83

Döring, Nicola, & M. Rohangis Mohseni. 2019. 'Male Dominance and Sexism on YouTube: Results of Three Content Analyses,' *Feminist Media Studies*, 19.4 (Taylor & Francis): 512–24

Dresner, Eli, & Susan C Herring. 2010. 'Functions of the Nonverbal in CMC: Emoticons and Illocutionary Force,' *Communication Theory*, 20.3: 249–68 <https://doi.org/10.1111/j.1468-2885.2010.01362.x>

Durham, Mercedes. 2007. 'Language Choice on a Swiss Mailing List,' in *The Multilingual Internet: Language, Culture, and Communication Online*, ed. by Brenda Danet & Susan C Herring (Oxford: Oxford University Press), pp. 319–39

Durrell, Martin. 2003. *Using German: A Guide to Contemporary Usage*, 2nd ed. (Cambridge: Cambridge University Press)

———. 2017. *Hammer's German Grammar and Usage*, 6th edn (London: Routledge) [accessed 22 September 2020]

Dürscheid, Christa. 2003. 'Medienkommunikation im Kontinuum von Mündlichkeit und Schriftlichkeit. Theoretische und empirische Proble- me,' *Zeitschrift für Angewandte Linguistik*: 37–56

———. 2005. 'Email - Verändert Sie Das Schreiben?' in *Websprache.net: Sprache Und Kommunikation Im Internet*, ed. by Torsten Siever, Peter Schlobinski, & Jens Runkehl (Berlin: Walter de Gruyter), pp. 85–97

———. 2006. 'Medienkommunikation Und Jugendsprache,' in *Perspektiven Der Jugendsprachforschung Trends and Developments in Youth Language Research*, ed. by Christa Dürscheid & Jürgen Spitzmüller (Frankfurt am Main: Peter Lang), pp. 117–31

———. 2016a. 'Graphematische Mikrovariation,' in *Handbuch Laut, Gebärde, Buchstabe*, ed. by Ulrike Domahs & Beatrice Primus (Berlin: De Gruyter), pp. 492–510

———. 2016b. 'Nähe, Distanz und neue Medien,' in *Zur Karriere von ›Nähe und Distanz‹*, ed. by Helmuth Feilke & Mathilde Hennig (Berlin: De Gruyter), pp. 357–86 <https://doi.org/10.1515/9783110464061-013>

Dürscheid, Christa, & Sarah Brommer. 2013. 'Ist Ein Freund Noch Ein Freund? Facebook Und Sprachwandel,' *Der Deutschunterricht*, 2.13: 28–40

———. 2016. 'Getippte Dialoge in Neuen Medien. Sprachkritische Aspekte Und Linguistische Analysen,' *Linguistik Online*, 37: 9

Dürscheid, Christa, & Karina Frick. 2014. 'Keyboard-to-Screen-Kommunikation Gestern Und Heute: SMS Und WhatsApp Im Vergleich,' in *Sprachen? Vielfalt! Sprache Und Kommunikation in Der Gesellschaft Und Den Medien. Eine Online-Festschrift Zum Jubiläum von Peter Schlobinski*, ed. by Alexa Mathias, Jens Runkehl, & Torsten Siever (Hannover: NetWorx 64), pp. 149–81

Dürscheid, Christa, & Christina Margrit Siever. 2017. 'Jenseits Des Alphabets – Kommunikation Mit Emojis,' *Zeitschrift für Germanistische Linguistik*, 45.2 (De Gruyter): 256–85 <https://doi.org/10.1515/zgl-2017-0013>

Dürscheid, Christa, Franc Wagner, & Sarah Brommer. 2010. *Wie Jugendliche schreiben: Schreibkompetenz und neue Medien* (Walter de Gruyter)

Eckert, Penelope. 2012. 'Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation,' *Annual Review of Anthropology*, 41: 87–100

Eichinger, Ludwig M. 2001. 'Spricht Die Jugend von Heute Das Deutsch von Morgen? Soziolinguistische Überlegungen Zum Normenwandel,' *Text & Kontext*, 23.2: 332–52

Elleström, Lars. 2016. 'Visual Iconicity in Poetry,' *Orbis Litterarum*, 71.6: 437–72 <https://doi.org/https://doi.org/10.1111/oli.12112>

Elsen, Hilke. 2009. 'Affixoide: Nur Was Benannt Wird, Kann Auch Verstanden Werden,' in *Deutsche Sprache: Zeitschrift für Theorie, Praxis, Dokumentation*, ed. by Ludwig M. Eichinger & Martine Dalmas, pp. 316–33

Emojipedia. 2017. 'Emoji Version 2.0 List' <https://emojipedia.org/emoji-2.0/> [accessed 24 September 2021]

———. 2021. '☑ Emoji Statistics [Updated September 2021]' <https://emojipedia.org/stats/> [accessed 1 December 2021]

Erickson, Thomas. 1999. 'Persistant Conversation: An Introduction,' *Journal of Computer-Mediated Communication*, 4.4 <https://doi.org/10.1111/j.1083-6101.1999.tb00105.x>

Evans, Vyvyan. 2017. *The Emoji Code. How Smiley Faces, Love Hearts and Thumbs Up are Changing the Way We Communicate* (London: Michael O'Mara Books Limited)

Fagan, Sarah M. B. 2009. *German: A Linguistic Introduction* (Cambridge: Cambridge University Press)

Felder, Samuel. 2020. 'Individuelle Verschriftungsmuster in Schweizerdeutschen WhatsApp-Chats,' in *Register Des Graphischen*, ed. by Jannis Androutsopoulos & Florian Busch (De Gruyter), pp. 93–132 <https://doi.org/10.1515/9783110673241-005>

Ferrara, Kathleen, Hans Brunner, & Greg Whittemore. 1991. 'Interactive Written Discourse as an Emergent Register,' *Written Communication*, 8.1: 8–34 <https://doi.org/10.1177/0741088391008001002>

Fladrich, Marcel, & Wolfgang Imo. 2020. '♀ J = ♂ J? Oder: Das Gelächter der Geschlechter 2.0: Emojigebrauch in der WhatsApp-Kommunikation,' in *Deutsch in Sozialen Medien: Interaktiv – multimodal – vielfältig*, ed. by Konstanze Marx, Henning Lobin, & Axel Schmidt (Berlin: De Gruyter), pp. 95–122 <https://doi.org/10.1515/9783110679885-006>

Flesch, Marie. 2019. '"That Spelling Tho": A Sociolinguistic Study of the Nonstandard Form of Though in a Corpus of Reddit Comments,' *European Journal of Applied Linguistics*, 7.2 (De Gruyter Oldenbourg): 163–88

Földes, Csaba. 2019. '„Jugendwort Des Jahres": Kommunikative Realität Oder Mediale Konstruktion von Jugendsprache?' *Wirkendes Wort*, 1.69: 121–39

Frehner, Carmen. 2008. *Email, SMS, MMS: The Linguistic Creativity of Asynchronous Discourse in the New Media Age* (Bern: Peter Lang)

Frey, Werner, & Federica Masiero. 2018. 'Desintegration Versus Parordination Bei "Obwohl"- Und "Weil"- Konstruktionen,' in *Im Mittelpunkt Deutsch*, 59th edn, ed. by André Meinunger (Berlin: Leibniz-Zentrum Allgemeine Sprachwissenschaft), pp. 57–82 <http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/46783>

Freywald, Ulrike. 2009. 'Kontexte für Nicht-Kanonische Verbzweitstellung: V2 Nach Dass Und Verwandtes,' in *Koordination Und Subordination Im Deutschen*, ed. by Veronika Ehrich et al. (Hamburg: Buske), pp. 113–34

———. 2010. '"Obwohl Vielleicht War Es Ganz Anders." Vorüberlegungen Zum Alter Der Verbzweitstellung Nach Subordinierenden Konjunktionen,' in *Historische Texgrammatik Und Historische Syntax Des Deutschen*, ed. by Arne Ziegler, pp. 54–84

———. 2016. 'V2-Nebensätze" – Ein Eigener Satztyp?' in *Satztypen Und Konstruktionen*, ed. by Rita Finkbeiner & Jörg Meibauer (Berlin: De Gruyter), pp. 326–72

Freywald, Ulrike, & Horst J Simon. 2007. 'Wenn die Wortbildung die Syntax stört: Über Verben, die nicht in V2 stehen können,' in *Verbale Wortbildung im Spannungsfeld zwischen Wort- semantik, Syntax und Rechtschreibung*, ed. by Maurice Kauffer & René Métrich (Tübingen: Stauffenburg), pp. 181–94

Funke, Wolfgang. 1982. 'Charakteristika Des Generationssoziolekts Der Jugendlichen in Der DDR,' *Kwartalnik Neofilologiczny*: 79–91

García, Ofelia. 2010. 'Languaging and Ethnifying,' in *Handbook of Language and Ethnic Identity*, ed. by Joshua A. Fishman & Ofelia García (Oxford: Oxford University Press), pp. 519–34

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

Garley, Matt. 2019. 'Choutouts: Language Contact and US-Latin Hip Hop on

YouTube,' *Lengua y Migración / Language and Migration*, 2.11: 77–106

Gaumann, Ulrike. 1983. '"Weil Die Machen Jetzt Bald Zu": Angabe-Und Junktivsatz in Der Deutschen Gegenwartssprache' (Universität Mainz)

Gawne, Lauren, & Gretchen McCulloch. 2019. 'Emoji as Digital Gestures,' *Language@Internet*, 17

Ge, Jing, & Ulrike Gretzel. 2018. 'Emoji Rhetoric: A Social Media Influencer Perspective,' *Journal of Marketing Management*, 34.15-16: 1272–95 <https://doi.org/10.1080/0267257X.2018.1483960>

Georgakopoulou, Alexandra, & Tereza Spilioti (eds.). 2016. *The Routledge Handbook of Language and Digital Communication*, Routledge Handbooks in Applied Linguistics (Abingdon: Routledge)

Geyken, Alexander. 2010. 'Statistical Variations of German Support Verb Constructions in Very Large Corpora,' *A Way with Words*: 1–13

Geyken, Alexander et al. 2017. 'Die Korpusplattform des „Digitalen Wörterbuchs der deutschen Sprache" (DWDS),' *Zeitschrift für germanistische Linguistik*, 45.2 <https://doi.org/10.1515/zgl-2017-0017>

Giannoulis, Elena, & Lukas R. A. Wilde (eds.). 2019. *Emoticons, Kaomoji, and Emoji: The Transformation of Communication in the Digital Age*, Routledge Research in Language and Communication, 6 (New York: Routledge)

Gibson, Will. 2008. 'Intercultural Communication Online: Conversation Analysis and the Investigation of Asynchronous Written Discourse,' *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 10

Glaznieks, Aivars, & Jennifer-Carmen Frey. 2020. 'Das DiDi-Korpus: Internetbasierte Kommunikation Aus Südtirol,' in *Deutsch in Sozialen Medien: Interaktiv – Multimodal – Vielfältig*, ed. by Konstanze Marx, Henning Lobin, & Axel Schmidt (Berlin: De Gruyter), pp. 353–54

Glaznieks, Alvars, & Jennifer-Carmen Frey. 2018. 'Dialekt als Norm? Zum Sprachgebrauch Südtiroler Jugendlicher auf Facebook,' in *Jugendsprachen: Aktuelle Perspektiven Internationaler Forschung*, ed. by Arne Ziegler (Berlin:

De Gruyter), pp. 859–90

Goffman, Erving. 1974. *Frame Analysis: An Essay on the Organization of Experience*. (Cambridge, MA: Harvard University Press)

———. 1981. *Forms of Talk* (Philadelphia: University of Pennsylvania Press)

Gohl, Christine, & Susanne Günthner. 1999. 'Grammatikalisierung von Weil Als Diskursmarker in Der Gesprochenen Sprache,' *Zeitschrift für Sprachwissenschaft*, 18.1: 39–75 <https://doi.org/10.1515/zfsw.1999.18.1.39>

Görke, Adrian. 2018. 'Jugendsprachliche Merkmale als Ausdruck glokaler jugendsprachlicher Tendenzen,' in *Jugendsprachen: Aktuelle Perspektiven Internationaler Forschung*, ed. by Arne Ziegler (Berlin: De Gruyter), pp. 97–122

Grice, H. Paul. 1975. 'Logic and Conversation,' in *Syntax and Semantics Volume 3: Speech Acts*, ed. by Peter Cole & Jerry Morgan (New York: Academic Press), pp. 41–58

Gruber, Helmut. 1997. 'Themenentwicklung in wissenschaftlichen E-mail-Diskussionslisten. Ein Vergleich zwischen einer moderierten und einer nichtmoderierten Liste,' in *Sprachwandel durch Computer.*, ed. by Rüdiger Weingarten (Opladen: Westdeutscher Verlag), pp. 105–28

Gruzd, Anatoliy, Barry Wellman, & Yuri Takhteyev. 2011. 'Imagining Twitter as an Imagined Community,' *American Behavioral Scientist*, 55.10 (Sage Publications Sage CA: Los Angeles, CA): 1294–1318

Gumperz, John J. 2009. 'The Speech Community,' in *Linguistic Anthropology: A Reader*, 2nd edn, ed. by Alessandro Duranti (Oxford: Wiley-Blackwell), pp. 66–73

Günther, Ulla, & Eva Lia Wyss. 1996. 'E-Mail-Briefe-Eine Neue Textsorte Zwischen Mündlichkeit Und Schriftlichkeit,' in *Textstrukturen Im Medienwandel*, ed. by Ernst Hess-Lüttich, Werner Holly, & Ulrich Püschel (Frankfurt am Main: Peter Lang), p. 61ß86

Günthner, Susanne. 1993. '«... Weil—man Kann Es Ja Wissenschaftlich Untersuchen». Diskurspragmatische Aspekte Der Wortstellung in WEIL-Sätzen,' *Linguistische Berichte*: 37–59

———. 1996. 'From Subordination to Coordination? Verb-Second Position in German Causal and Concessive Constructions,' *Pragmatics*, 6.3 (John Benjamins): 323–56

———. 1999. 'Entwickelt Sich Der Konzessivkonnektor Obwohl Zum Diskursmarker? Grammatikalisierungstendenzen Im Gesprochenen Deutsch,' *Linguistische Berichte*: 409–46

———. 2000a. 'Sprechen Wir Ungrammatisch? Zur Verwendung von Weil Und Obwohl Mit Hauptsatzstellung Im Gesprochenen Deutsch,' in *Germanistentreffen Deutschland-Indien-Indonesien-Philippinen-Taiwan-Thailand-Vietnam1999*, ed. by Deutsche Akademische Austausch Dienst (Bonn: Rosch), pp. 243–60

———. 2000b. '"wobei (.) es hat alles immer zwei seiten." Zur Verwendung von wobei im gesprochenen Deutsch,' *Deutsche Sprache*: 313–41

———. 2002. 'Konnektoren Im Gesprochenen Deutsch: Normverstoß Oder Funktionale Differenzierung?' *Deutsch Als Fremdsprache*, 39.2: 67–74

Günthner, Susanne, & Wolfgang Imo. 2003. 'Die Reanalyse von Matrixsätzen Als Diskursmarker: Ich Mein-Konstruktionen Im Gesprochenen Deutsch': 35

Haase, Martin et al. 1997. 'Internetkommunikation Und Sprachwandel,' in *Sprachwandel Durch Computer*, ed. by Rüdiger Weingarten (Opladen: Westdeutscher Verlag), pp. 51–85

Hamilton, William A., Oliver Garretson, & Andruid Kerne. 2014. 'Streaming on Twitch: Fostering Participatory Communities of Play Within Live Mixed Media,' in *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems* (Toronto: ACM), pp. 1315–24

Haryanto, Nurul, & Kandi Aryani Suwito. 2020. 'Gender Identity Construction

of Male Beauty Vloggers on YouTube,' *Journal of Talent Development and Excellence*, 12.2s: 171–77

Heinemann, Margot. 1979. 'Bemerkungen Zur Sprechweise Der Jugendlichen in Der DDR,' *Sprachpflege*, 28: 116–18

Hellinger, Marlis. 1995. 'Language and Gender,' in *The German Language in the Real World*, ed. by Patrick Stevenson (New York: Oxford University Press), pp. 281–316

Henne, Helmut. 1981. 'Jugend und ihre Sprache,' in *Jahrbuch des Instituts für deutsche Sprache 1980*, ed. by Peter Schröder & Hugo Steger (Düsseldorf: Institut für deutsche Sprache), pp. 370–84

———. 1986. *Jugend und ihre Sprache: Darstellung, Materialien, Kritik*, 1st edn (New York: De Gruyter)

Hentschel, Elke. 1998. 'Communication on IRC,' *Linguistik Online* <https://bop.unibe.ch/linguistik-online/article/download/1084/1773> [accessed 24 January 2020]

Heritage, Frazer, & Veronika Koller. 2020. 'Incels, in-Groups, and Ideologies: The Representation of Gendered Social Actors in a Sexuality-Based Online Community,' *Journal of Language and Sexuality*, 9.2: 152–78 <https://doi.org/10.1075/jls.19014.her>

Herring, Susan C. 1992. 'Gender and Participation in Computer-Mediated Linguistic Discourse.' in *Conference of the Annual Meeting of the Linguistic Society of America* (Philadelphia: ERIC)

——— (ed.). 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: Allyn & Bacon), pp. 144–54

———. 1999. 'Interactional Coherence in CMC,' *Journal of Computer-Mediated Communication*, 4 <https://doi.org/10.1111/j.1083-6101.1999.tb00106.x>

———. 2007. 'A Faceted Classification Scheme for Computer-Mediated Discourse,' *Language@Internet*, 4.1 <http://www.languageatinternet.org/articles/2007/761>

———. 2008. 'Questioning the Generational Divide: Technological Exoticism and Adult Constructions of Online Youth Identity,' *Youth, Identity, and Digital Media*: 71–94

———. 2012. 'Grammar and Electronic Communication,' in *The Encyclopedia of Applied Linguistics*, ed. by Carol A Chapelle (Oxford: Blackwell), p. wbeal0466 <https://doi.org/10.1002/9781405198431.wbeal0466>

———. 2013. 'Discourse in Web 2.0: Familiar, Reconfigured, and Emergent,' in *Discourse 2.0: Language and New Media*, ed. by Deborah Tannen & Anne Marie Trester (Washington D.C.: Georgetown University Press), pp. 1–26

———. 2019. 'The Coevolution of Computer-Mediated Communication and Computer-Mediated Discourse Analysis,' in *Analyzing Digital Discourse: New Insights and Future Directions*, ed. by Patricia Bou-Franch & Pilar Garcés-Conejos Blitvich (Cham: Palgrave Macmillan), pp. 25–67

Herring, Susan C et al. 2020a. 'Animoji Adoption and Use: Gender Associations with an Emergent Technology,' p. 10

Herring, Susan C et al. 2020b. 'Animoji Performances: "Cuz I Can Be a Sexy Poop",' *Language@Internet*, 18.1 <https://www.languageatinternet.org/articles/2020/herring>

Herring, Susan C, & Ashley Dainas. 2017. '"Nice Picture Comment!" Graphicons in Facebook Comment Threads,' in *Proceedings of the 50th Hawaii International Conference on System Sciences* (Los Alamitos) <http://hdl.handle.net/10125/41419> [accessed 21 June 2019]

Herring, Susan C, & Ashley R Dainas. 2020. 'Gender and Age Influences on Interpretation of Emoji Functions,' *ACM Transactions on Social Computing*, 3.2: 10:1–26 <https://doi.org/10.1145/3375629>

Herring, Susan C, Dieter Stein, & Tuija Virtanen (eds.). 2013. *Pragmatics of Computer-Mediated Communication* (Berlin: De Gruyter)

Hess, Wolfgang J., Klaus J. Kohler, & Hans-Günther Tillmann. 1995. 'The Phondat-Verbmobil Speech Corpus,' in *Fourth European Conference on Speech Communication and Technology*

Hill Collins, Patricia. 2002. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*, 2nd edn (London: Taylor and Francis)

Hilte, Lisa, Walter Daelemans, & Reinhild Vandekerckhove. 2020. 'Lexical Patterns in Adolescents' Online Writing: The Impact of Age, Gender, and Education,' *Written Communication*, 37.3 (SAGE Publications Inc): 365–400 <https://doi.org/10.1177/0741088320917921>

Hilte, Lisa, Reinhild Vandekerckhove, & Walter Daelemans. 2019. 'Expressive Markers in Online Teenage Talk,' *Nederlandse Taalkunde*, 23.3: 293–323 <https://doi.org/10.5117/NEDTAA2018.3.003.HILT>

Hiltz, Starr Roxanne, & Murray Turoff. 1993. *The Network Nation: Human Communication via Computer*, 2nd edn (Boston: MIT Press)

Hinrichs, Lars. 2018. 'The Language of Diasporic Blogs,' in *Multilingual Youth Practices in Computer Mediated Communication*, ed. by Cecelia Cutler & Unn Røyneland (Cambridge: Cambridge University Press), pp. 186–204 <https://doi.org/10.1017/9781316135570.011>

Hinz, Leonore. 2015. 'The Functions of Emoticons and Pictograms in Instant Messengers': 12

Hofmann, Ute. 2018. 'Interaktion von Sprachwandel und Sprachvarietäten,' in *Jugendsprachen: Aktuelle Perspektiven Internationaler Forschung*, ed. by Arne Ziegler (Berlin: De Gruyter), pp. 67–83

Holmes, Janet. 1984. 'Modifying Illocutionary Force,' *Journal of Pragmatics*,

8.3: 345–65 <https://doi.org/10.1016/0378-2166(84)90028-6>

———. 2013. *An Introduction to Sociolinguistics*, 4th edn (London: Routledge)

Holmes, Janet, & Miriam Meyerhoff (eds.). 2003. *The Handbook of Language and Gender* (Oxford: John Wiley & Sons)

Holtgraves, Thomas, & Caleb Robinson. 2020. 'Emoji Can Facilitate Recognition of Conveyed Indirect Meaning,' *PLOS ONE*, 15.4 (Public Library of Science): 1–13 <https://doi.org/10.1371/journal.pone.0232361>

Hougaard, Tina Thode, & Marianne Rathje. 2018. 'Emojis in the Digital Writings of Young Danes,' in *Jugendsprachen: Aktuelle Perspektiven Internationaler Forschung*, ed. by Arne Ziegler (Berlin: De Gruyter), pp. 773–806 <https://doi.org/10.1515/9783110472226-035>

Höflich, Joachim R. 2003. 'Vermittlungskulturen Im Wandel: Brief, E-Mail, SMS,' in *Vermittlungskulturen Im Wandel: Brief, E-Mail, SMS*, ed. by Joachim R. Höflich & Julian Gebhardt (Frankfurt am Main: Peter Lang), pp. 39–62

Huffaker, David A., & Sandra L. Calvert. 2005. 'Gender, Identity, and Language Use in Teenage Blogs,' *Journal of Computer-Mediated Communication*, 10 <https://doi.org/10.1111/j.1083-6101.2005.tb00238.x>

Hugger, Kai-Uwe. 2014. 'Digitale Jugendkulturen: Eine Einleitung,' in *Digitale Jugendkulturen*, ed. by Kai-Uwe Hugger (Wiesbaden: VS Verlag für Sozialwissenschaften), pp. 11–28

Hutchby, Ian, & Robin Wooffitt. 1998. *Conversation Analysis: Principles, Practices, and Applications* (Cambridge: Polity Press)

IFPI. 2017. *Connecting with Music: Music Consumer Insight Report* (London: International Federation of the Phonographic Industry) <https://ifpi.org/downloads/Music-Consumer-Insight-Report-2017.pdf> [accessed 22 February 2019]

Ilbury, Christian. 2020. '"Sassy Queens": Stylistic Orthographic Variation in Twitter and the Enregisterment of AAVE,' *Journal of Sociolinguistics*, 24.2: 245–64 <https://doi.org/10.1111/josl.12366>

Imo, Wolfgang. 2011. 'Clines of Subordination: Constructions with the German 'Complement-Taking Predicate'glauben,' in *Subordination in Conversation: A Cross-Linguistic Perspective*, pp. 165–90

Ingram, Matthew Bruce. 2019. 'YouTube Commentaries on Trans Time-Lapse Videos: Transforming Misgendering Stances into Pedagogical Moments,' *Somatechnics*, 9.1: 32–57 <https://doi.org/10.3366/soma.2019.0264>

Jacobson, Robert E., & Robert Jacobson. 1999. *Information Design* (MIT press)

Jaeger, Sara R. et al. 2018. 'Emoji Questionnaires Can Be Used with a Range of Population Segments: Findings Relating to Age, Gender and Frequency of Emoji/Emoticon Use,' *Food Quality and Preference*, 68: 397–410 <https://doi.org/10.1016/j.foodqual.2017.12.011>

Jones, Graham M., Bambi B. Schieffelin, & Rachel E. Smith. 2011. 'When Friends Who Talk Together Stalk Together: Online Gossip as Metacommunication,' in *Digital Discourse: Language in the New Media*, ed. by Crispin Thurlow & Kristine R. Mroczek (Oxford: Oxford University Press), pp. 26–47

Jones, Lucy. 2012. *Dyke/Girl: Language and Identities in a Lesbian Group* (London: Palgrave Macmillan) <https://doi.org/10.1057/9781137271341>

Jones, Rodney H., & Christoph A. Hafner. 2012. *Understanding Digital Literacies: A Practical Introduction* (Abingdon: Routledge)

Jones, Steve (ed.). 1998. *Cybersociety 2.0: Revisiting Computer-Mediated Communication Community* (Thousand Oaks: SAGE), II

Jucker, Andreas H., & Christa Dürscheid. 2012. 'The linguistics of keyboard-to-screen communication. A new terminological framework,' *Linguistik Online*, 56.6/12 (European University Viadrina): 1–26 <https://doi.org/10.5167/uzh-67310>

Jurčević, Jana. 2019. 'Sound Symbolism of Contemporary Communication: Croatian Imitative Expressions in Computer-Mediated Discourse,' in *Language, Music and Computing*, ed. by Polina Eismont, Olga Mitrenina, &

Asya Pereltsvaig (Cham: Springer), CMXLIII, pp. 91–110 <https://doi.org/10.1007/978-3-030-05594-3_8>

Kallmeyer, Werner, & Inken Keim. 2003. 'Linguistic Variation and the Construction of Social Identity in a GermanTurkish Setting: A Case Study of an Immigrant Youth Group in Mannheim, Germany,' in *Discourse Constructions of Youth Identities*, ed. by Jannis Androutsopoulos & Alexandra Georgakopoulou (Philadelphia: John Benjamins), pp. 29–46

Kantrowitz, Barbara, & D. Rosenberg. 1994. 'Men, Women & Computers.' *Newsweek*, 123.20 (Newsweek): 48–55

Kavanagh, Barry. 2019. 'Exploration of Kaomoji, Emoji, and Kigō,' in *Emoticons, Kaomoji, and Emoji: The Transformation of Communication in the Digital Age*, Routledge Research in Language and Communication, 6, ed. by Elena Giannoulis & Lukas R. A. Wilde (New York: Routledge), pp. 148–67

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

Keim, Inken, & Ralf Knöbl. 2007. 'Sprachliche Varianz Und Sprachliche Virtuosität türkisch-Stämmiger Ghetto-Jugendlicher in Mannheim,' in *Standard, Variation Und Sprachwandel in Germanischen Sprachen*, Studien Zur Deutschen Sprache, 41 (Tübingen: Narr), pp. 157–99

Keller, Rudi. 1995. 'The Epistemic Weil,' in *Subjectivity and Subjectivisation: Linguistic Perspectives*, ed. by Dieter Stein & Susan Wright (Cambridge: Cambridge University Press), pp. 16–30 <https://www.cambridge.org/core/books/subjectivity-and-subjectivisation/epistemic-weil/4B86ED9768CCF0CF9659B8FDA685BF33>

Kelly, Ryan, & Leon Watts. 2015. 'Characterising the Inventive Appropriation of Emoji as Relationally Meaningful in Mediated Close Personal Relationships,' in *Experiences of Technology Appropriation: Unanticipated Users, Usage, Circumstances, and Design* (Oslo), xx

Kempen, Gerard, & Karin Harbusch. 2016. 'Verb-Second Word Order After German Weil "Because": Psycholinguistic Theory from Corpus-Linguistic Data,' *Glossa: A Journal of General Linguistics*, 1.1: 3 <https://doi.org/10.5334/gjgl.46>

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

Kennedy, Graeme D. 1998. *An Introduction to Corpus Linguistics* (London: Longman)

Kessler, Florence. 2008. 'Instant Messaging: Eine Neue Interpersonale Kommunikationsform,' *NetWorx*

Kettunen, Kimmo. 2014. 'Can Type-Token Ratio Be Used to Show Morphological Complexity of Languages?' *Journal of Quantitative Linguistics*, 21.3: 223–45 <https://doi.org/10.1080/09296174.2014.911506>

Kiesler, Sara, Jane Siegel, & Timothy W. McGuire. 1984. 'Social Psychological Aspects of Computer-Mediated Communication,' *American Psychologist*, 39: 1123–34

Kilian, Jörg. 2001. 'T@stentöne. Geschriebene Umgangssprache in Computervermittelter Kommunikation. Historisch-Kritische Ergänzungen Zu Einem Neuen Feld Der Linguistischen Forschung,' in *Chat-Kommunikation. Sprache, Interaktion, Sozialität & Identität in Synchroner Computervermittelter Kommunikation*, ed. by Michael Beißwenger (Stuttgart: Ibidem), pp. 55–78

Kleinberger Günther, Ulla, & Carmen Spiegel. 2006. 'Jugendliche Schreiben Im Internet: Grammatische Und Orthographische Phänomene in Normungebundenen Kontexten,' in *Perspektiven Der*

*Jugendsprachforschung- Trends and Developments in Youth Language Research*, ed. by Christa Dürscheid & Jürgen Spitzmüller (Frankfurt am Main: Peter Lang), pp. 101–15

Kluge, Friedrich, & Elmar Seebold (eds.). 2002. *Etymologisches Wörterbuch der deutschen Sprache*, 24th edn (Berlin: Walter de Gruyter) [accessed 15 August 2021]

Know Your Meme. 2021. 'Mocking SpongeBob,' *Know Your Meme* <https://knowyourmeme.com/memes/mocking-spongebob> [accessed 9 April 2021]

Koch, Peter, & Wulf Österreicher. 1985. 'Sprache Der Nähe–Sprache Der Distanz,' *Romanistisches Jahrbuch*, 36: 15–43

———. 2007. 'Schriftlichkeit Und Kommunikative Distanz,' *Zeitschrift für Germanistische Linguistik*, 35: 346–75

Koch, Timo, Peter Romero, & Clemens Stachl. 2020. 'Predicting Age and Gender from Language, Emoji, and Emoticon Use in WhatsApp Instant Messages,' *PsyArXiv* <https://doi.org/10.1016/j.chb.2021.106990>

Kohler, Klaus J. 1995. *Einführung in Die Phonetik Des Deutschen*, 2nd edn (Berlin: E. Schmidt)

Kohler, Klaus J., & Jonathan Rodgers. 2001. 'Schwa Deletion in German Read and Spontaneous Speech,' *Spontaneous German Speech: Symbolic Structures and Gestural Dynamics*, 35: 97–123

Konrad, Artie, Susan C Herring, & David Choi. 2020. 'Sticker and Emoji Use in Facebook Messenger: Implications for Graphicon Change,' *Journal of Computer-Mediated Communication*, 25.3: 217–35 [accessed 4 February 2021]

Kotsinas, Ulla-Britt. 1997. 'Young Peoples Language. Norm, Variation and Language Change,' *Stockholm Studies in Modern Philology, New Series*, 11: 109–32

Kozinets, Robert. 2010. *Netnography: Doing Ethnographic Research Online*

(London: SAGE)

Könning, Benjamin. 2018. 'Peer-Kommunikation in der Schule: empirische und methodische Zugänge,' in *Jugendsprachen: Aktuelle Perspektiven Internationaler Forschung*, ed. by Arne Ziegler (Berlin: De Gruyter), pp. 247–68

———. 2020. 'Jugendsprache Und Jugendkommunikation,' in *Handbuch Sprachkritik*, ed. by Thomas Niehr, Jörg Kilian, & Jürgen Schiewe (Berlin: J. B. Metzler), pp. 333–42

Kunkel-Razum, Kathrin et al. (eds.). 2020. *Duden - die deutsche Rechtschreibung: auf der Grundlage der aktuellen amtlichen Rechtschreibregeln*, Der Duden: in zwölf Bänden; das Standardwerk zur deutschen Sprache, 28th edn (Berlin: Dudenverlag)

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

Landert, Daniela, & Andreas H. Jucker. 2011. 'Private and Public in Mass Media Communication: From Letters to the Editor to Online Commentaries,' *Journal of Pragmatics*, 43.5: 1422–34

Langenscheidt (ed.). 2014. *100% Jugendsprache 2014* (Munich: Langenscheidt)

——— (ed.). 2016. *100% Jugendsprache 2016* (Munich: Langenscheidt)

———. 2018. 'Jugendwort Des Jahres 2018' <https://www.langenscheidt.com/jugendwort-des-jahres> [accessed 18 January 2019]

LaViolette, Jack. 2017. 'Cyber-Metapragmatics and Alterity on Reddit.com,' *Working Papers in Urban Language & Literacies*, WP229: 1–15 <https://www.academia.edu/34945772/WP229_LaViolette_2017._Cyber-metapragmatics_and_alterity_on_reddit.com> [accessed 5 January 2019]

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

Lee, Carmen. 2016. *Multilingualism Online* (Oxford: Routledge)

Leibniz-Institut für Deutsche Sprache. 2017a. 'Koordination,' *Grammatisches Informationssystem, Grammis* <https://grammis.ids-mannheim.de/terminologie/1233> [accessed 17 May 2020]

———. 2017b. 'Wortstellung. In: "Systematische Grammatik",' *Grammatisches Informationssystem, Grammis* <https://grammis.ids-mannheim.de/systematische-grammatik/748> [accessed 18 May 2020]

———. 2018a. 'Postponierer. In: "Systematische Grammatik",' *Grammatisches Informationssystem, Grammis* <https://grammis.ids-mannheim.de/systematische-grammatik/1203> [accessed 18 May 2020]

———. 2018b. 'Rechtes Außenfeld,' *Grammatisches Informationssystem, Grammis* <https://grammis.ids-mannheim.de/systematische-grammatik/1137> [accessed 9 September 2021]

———. 2018c. 'Topikalisierung. In: "Wissenschaftliche Terminologie",' *Grammatisches Informationssystem, Grammis* <https://grammis.ids-mannheim.de/terminologie/598> [accessed 30 October 2020]

———. 2018d. 'Verbalkomplex. In: "Systematische Grammatik",' *Grammatisches Informationssystem, Grammis* <https://doi.org/10.14618/grammatiksystem>

———. 2018e. 'Wörterbuch Der Konnektoren,' *Grammatisches Informationssystem, Grammis* <https://doi.org/10.14618/wb-konnektoren>

Leuschner, Torsten, & Eline Wante. 2009. 'Personale Suffixoide Im Deutschen Und Niederländischen. Methode Und Ergebnisse Einer Vergleichenden Korpusstudie,' *Germanistische Mitteilungen*: 59–73

Levon, Erez. 2015. 'Integrating Intersectionality in Language, Gender, and Sexuality Research,' *Language and Linguistics Compass*, 9.7: 295–308 <https://doi.org/10.1111/lnc3.12147>

Liddicoat, Anthony. 2022. *An Introduction to Conversation Analysis*, 3rd edn (London: Bloomsbury Academic)

Ljubešić, Nikola, & Darja Fišer. 2016. 'A Global Analysis of Emoji Usage,' in *Proceedings of the 10th Web as Corpus Workshop* (Berlin: Association for Computational Linguistics), pp. 82–89 <https://doi.org/10.18653/v1/W16-2610>

Lo, Adrienne. 1999. 'Codeswitching, Speech Community Membership, and the Construction of Ethnic Identity,' *Journal of Sociolinguistics*, 3: 461–79 <https://doi.org/10.1111/1467-9481.00091>

Lu, Xuan et al. 2016. 'Learning from the Ubiquitous Language: An Empirical Analysis of Emoji Usage of Smartphone Users,' in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (ACM), pp. 770–80

Lüngen, Harald, & Marc Kupietz. 2020. 'IBK- Und Social Media-Korpora Am Leibniz-Institut für Deutsche Sprache,' in *Deutsch in Sozialen Medien: Interaktiv – Multimodal – Vielfältig*, ed. by Konstanze Marx, Henning Lobin, & Axel Schmidt (Berlin: De Gruyter), pp. 319–42 <https://doi.org/10.1515/9783110679885-016>

Lytra, Vally. 2016. 'Language and Ethnic Identity,' in *The Routledge Handbook of Language and Identity*, ed. by Siân Preece (Abingdon: Routledge), pp. 131–45

Mackenzie, Jai. 2018. *Language, Gender and Parenthood Online: Negotiating Motherhood in Mumsnet Talk* (London: Routledge)

Markowsky, George. 1992. 'Misconceptions about the Golden Ratio,' *The College Mathematics Journal*, 23.1: 2–19

Martín Zorraquino, María Antonia, & José Portolés. 1999. 'Los Marcadores Del

Discurso,' in *Gramática Descriptiva de La Lengua Española*, ed. by Ignacio Bosque & Violeta Demonte (Madrid: Espasa Calpe), III, pp. 4051–4213

Marwick, Alice E., & danah boyd. 2011. 'I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience,' *New Media & Society*, 13.1: 114–33 <https://doi.org/10.1177/1461444810365313>

Marx, Konstanze, & Georg Weidacher. 2020. *Internetlinguistik: Ein Lehr- und Arbeitsbuch* (Tübingen: Narr)

Massa, Eleonora. 2016. 'Über Das Schreibverhalten Jugendlicher in Und Außerhalb Der Schule: Eine Empirische Untersuchung,' in *Jugendsprache in Schule, Medien Und Alltag*, ed. by Carmen Spiegel & Daniel Gysin (Frankfurt am Main: Peter Lang), pp. 125–38 <https://www.peterlang.com/view/title/17633> [accessed 25 July 2021]

McAteer, Erica. 1992. 'Typeface Emphasis and Information Focus in Written Language,' *Applied Cognitive Psychology*, 6.4: 345–59 [accessed 5 March 2021]

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

McEnery, Tony, Richard Xiao, & Yukio Tono. 2006. *Corpus-Based Language Studies: An Advanced Resource Book* (Abingdon: Routledge)

Meer, Dorothee. 2018. '„Liebe Dagi Bee Du Bist Wunder Wunder hübsch" – Osmotische Werbung Und Jugendliche Identitätsarbeit Im Rahmen von Kommentarlisten Auf YouTube,' in *Jugendliche Im Gespräch* (Bern: Peter Lang), pp. 299–328

Mendoza-Denton, Norma. 2016. 'Norteño and Sureño Gangs, Hip Hop, and Ethnicity on YouTube: Localism in California Through Spanish Accent Variation,' in *Raciolinguistics: How Language Shapes Our Ideas about Race*, ed. by John R. Rickford (Oxford University Press), pp. 135–50

Merchant, Guy. 2001. 'Teenagers in Cyberspace: An Investigation of Language Use and Language Change in Internet Chatrooms,' *Journal of Research in*

*Reading*, 24.3: 293–306 <https://doi.org/https://doi.org/10.1111/1467-9817.00150>

Meredith, Joanne. 2019. 'Conversation Analysis and Online Interaction,' *Research on Language and Social Interaction*, 52.3: 241–56 <https://doi.org/10.1080/08351813.2019.1631040>

Mešic, Sanela. 2018. 'Fremdwörter in der Jugendsprache vor 20 Jahren und heute,' in *Jugendsprachen: Aktuelle Perspektiven Internationaler Forschung*, ed. by Arne Ziegler (Berlin: De Gruyter), pp. 443–54

Migge, Bettina. 2020. 'Mediating Creoles: Language Practices on a YouTube Show,' *Journal of Pidgin and Creole Languages*, 35.2 (John Benjamins): 381–404

Milani, Tommaso M. 2013. 'Are "Queers" Really "Queer?" Language, Identity and Same-Sex Desire in a South African Online Community,' *Discourse & Society*, 24.5: 615–33

Milani, Tommaso M., & Rickard Jonsson. 2011. 'Incomprehensible Language? Language, Ethnicity and Heterosexual Masculinity in a Swedish School,' *Gender and Language*, 5.2: 239–36

Miyake, Kazuko. 2007. 'How Young Japanese Express Their Emotions Visually in Mobile Phone Messages: A Sociolinguistic Analysis,' *Japanese Studies*, 27.1: 53–72 <https://doi.org/10.1080/10371390701268646>

Miyashita, Hiroyuki. 2003. 'Weil, Obwohl, während Und Wobei: Warum Werden Sie V2-Konjunktionen Und Nicht Andere,' *Energeia*, 28.2003: 59–81

Moore, David S., George P. McCabe, & Bruce A. Craig. 2017. *Introduction to the Practice of Statistics*, 9th edn (New York: Macmillan Education)

Moschini, Ilaria. 2016. 'The "Face with Tears of Joy" Emoji. A Socio-Semiotic and Multimodal Insight into a Japan-America Mash-Up,' *HERMES - Journal of Language and Communication in Business*: 11–25 <https://doi.org/10.7146/hjlcb.v0i55.24286>

Mostovaia, Irina. 2018. 'Nonverbale graphische Ressourcen bei Reparaturen in

der interaktionalen informellen Schriftlichkeit am Beispiel der deutschen Chat-Kommunikation via IRC-Chat und WhatsApp,' *Journal für Medienlinguistik*, 1.1: 42–79 <https://doi.org/10.21248/jfml.2018.6>

Mroczynski, Robert. 2012. *Grammatikalisierung und Pragmatikalisierung: zur Herausbildung der Diskursmarker wobei, weil und ja im gesprochenen Deutsch*, Tübinger Beiträge zur Linguistik, 530 (Tübingen: Narr)

———. 2018. 'Geil oder doch porno? Zum Gebrauch und Wandel von emphatisch-evaluativen Ausdrücken in der Jugendsprache,' in *Jugendsprachen: Aktuelle Perspektiven Internationaler Forschung*, ed. by Arne Ziegler (Berlin: De Gruyter), pp. 325–42

Müller, Rolf. 1971. 'Die Merkmale für "Abhängigkeit" Bei Uneingeleiteten Gliedsätzen in Transkriptionen Gesprochener Texte,' in *Forschungen Zur Gesprochenen Sprache Und Möglichkeiten Ihrer Didaktisierung*, ed. by Goethe Institute (Munich: Goethe Institute), pp. 119–25

Müller, Stefan. 2003. 'Mehrfache Vorfeldbesetzung,' *Deutsche Sprache*, 31.1: 29–62

———. 2019. *Grammatical Theory: From Transformational Grammar to Constraint-Based Approaches*, 3rd edn (Berlin: Language Science Press) <https://doi.org/10.26530/OAPEN_611693>

Na'aman, Noa, Hannah Provenza, & Orion Montoya. 2017. 'Varying Linguistic Purposes of Emoji in (Twitter) Context,' in *Proceedings of ACL 2017, Student Research Workshop* (Vancouver: Association for Computational Linguistics), pp. 136–41 <https://www.aclweb.org/anthology/P17-3022> [accessed 8 December 2020]

Nakamura, Lisa. 2002. *Cybertypes: Race, Ethnicity, and Identity on the Internet* (Florence: Taylor and Francis)

Nastri, Jacqueline, Jorge Pena, & Jeffrey T. Hancock. 2006. 'The Construction of Away Messages: A Speech Act Analysis,' *Journal of Computer-Mediated Communication*, 11: 1025–45

Neuland, Eva. 1986. 'Jugendsprache Im Gesellschaftlichen Wandel,' *Der Deutschunterricht*, 4: 52–75

———. 1987. 'Spiegelungen Und Gegenspiegelungen. Anregungen für Eine Zukünftige Jugendsprachenforschung./Reflections and Counter-Reflections. Suggestions for Future Youth Language Research,' *Zeitschrift für Germanistische Linguistik*, 15: 58–82 [accessed 6 July 2021]

———. 1994. 'Jugendsprache Und Standardsprache. Zum Wechselverhältnis von Stilwandel Und Sprachwandel,' *Zeitschrift für Germanistik*, 4.1: 78–98 [accessed 19 May 2021]

———. 2003. 'Doing Youth: Zur medialen Konstruktion von Jugend und Jugendsprache,' in *Jugendsprache - Jugendliteratur - Jugendkultur*, ed. by Eva Neuland (Frankfurt am Main: Peter Lang), pp. 261–74

———. 2006. 'Deutsche Schülersprache Heute: Total Normal?' in *Perspektiven Der Jugendsprachforschung- Trends and Developments in Youth Language Research*, ed. by Christa Dürscheid & Jürgen Spitzmüller (Frankfurt am Main: Peter Lang), pp. 51–72

———. 2018. *Jugendsprache: Eine Einführung*, 2nd edn (Tübingen: Narr)

Neuland, Eva, Benjamin Könning, & Elisa Wessels. 2018. 'Gebrauchs- Und Verständnisweisen Sprachlicher Höflichkeit von Jugendlichen: Korrespondenzanalysen von Fragebogen- Und Spontandaten Am Beispiel von Komplimentieren/Loben,' in *Jugendliche Im Gespräch* (Bern: Peter Lang), pp. 15–38

Newgrounds. 2012. 'Youtube Hates the Enter Key.' *Newgrounds.com* <https://www.newgrounds.com/bbs/topic/1309016> [accessed 10 March 2021]

Novak, Petra Kralj et al. 2015. 'Sentiment of Emojis,' *PLOS ONE*, 10.12 (Public Library of Science): e0144296 <https://doi.org/10.1371/journal.pone.0144296>

Nowotny, Andrea. 2005. 'Daumenbotschaften. Die Bedeutung von Handy

Und SMS für Jugendliche,' *NetWorx* <http://www.mediensprache.net/networx/networx-44.pdf> [accessed 26 July 2021]

O'Reilly, Tim, & Media O'Reilly. 2005. 'What is Web 2.0?: design patterns and business models for the next generation of software' <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>

Oberdorfer, Georg, & Anna Weiß. 2018. 'Youth Language(s) in Austria - State of Research and First Findings,' in *Jugendsprachen: Aktuelle Perspektiven Internationaler Forschung*, ed. by Arne Ziegler (Berlin: De Gruyter), pp. 463–88

Omoniyi, Tope. 2016. 'Language, Race and Identity,' in *The Routledge Handbook of Language and Identity*, ed. by Siân Preece (Abingdon: Routledge), pp. 146–62

Ong, Walter J. 1982. *Orality and Literacy* (Abingdon: Routledge)

Onysko, Alexander. 2007. *Anglicisms in German: Borrowing, Lexical Productivity, and Written Codeswitching* (Berlin: Walter de Gruyter)

Padilla. 1999. 'Psychology,' in *Handbook of Language and Ethnic Identity*, 1st edn, ed. by Joshua A. Fishman (Oxford: Oxford University Press), pp. 109–21

Panizo-LLedot, Ańgel et al. 2019. 'Describing Alt-Right Communities and Their Discourse on Twitter During the 2018 Us Mid-Term Elections,' in *Complex Networks and Their Applications VIII*, ed. by Hocine Cherifi et al. (Cham: Springer), pp. 427–39

Paolillo, John. 2011. '"Conversational" Codeswitching on Usenet and Internet Relay Chat,' *Language@Internet*, 8.3: 1–2

Papacharissi, Zizi. 2004. 'Democracy Online: Civility, Politeness, and the Democratic Potential of Online Political Discussion Groups,' 6: 259–83 <https://doi.org/10.1177/1461444804041444>

——— (ed.). 2011. *A Networked Self: Identity, Community and Culture on Social Network Sites* (New York: Routledge)

———. 2012. 'Without You, I'm Nothing: Performances of the Self on Twitter,' *International Journal of Communication*, 6: 18

Papafragou, Anna, Kimberly Cassidy, & Lila Gleitman. 2007. 'When We Think about Thinking: The Acquisition of Belief Verbs,' *Cognition*, 105.1: 125–65 <https://doi.org/10.1016/j.cognition.2006.09.008>

Papasavva, Antonis et al. 2020. 'Raiders of the Lost Kek: 3.5 Years of Augmented 4chan Posts from the Politically Incorrect Board,' *Proceedings of the International AAAI Conference on Web and Social Media*, 14: 885–94 <https://ojs.aaai.org/index.php/ICWSM/article/view/7354> [accessed 6 April 2021]

Parkins, Róisín. 2012. 'Gender and Emotional Expressiveness: An Analysis of Prosodic Features in Emotional Expression,' *Griffith Working Papers in Pragmatics and Intercultural Communication*, 1.5: 46–54

Pasch, Renate et al. 2003. *Handbuch der deutschen Konnektoren: linguistische Grundlagen der Beschreibung und syntaktische Merkmale der deutschen Satzverknüpfer (Konjunktionen, Satzadverbien und Partikeln)* (Berlin: De Gruyter)

Pauli, Stephanie. 2011. *'Ey Alter, Du Bist Voll Der Wort-Checker!': Jugendsprache: Eine Empirische Untersuchung Der Spracheinstellungen von Jugendlichen Und Erwachsenen* (Hamburg: tredition) <https://books.google.co.uk/books?id=arNwN78xykMC>

Pavalanathan, Umashanthi, & Jacob Eisenstein. 2015. 'Emoticons Vs. Emojis on Twitter: A Causal Inference Approach,' *CoRR*, 1510.08480

———. 2016. 'More Emojis, Less :) The Competition for Paralinguistic Function in Microblog Writing,' *First Monday* <https://doi.org/10.5210/fm.v21i11.6879>

Pennycook, Alastair, & Emi Otsuji. 2015. *Metrolingualism: Language in the City*

(London: Routledge)

Petitjean, Cécile, & Etienne Morel. 2017. '"Hahaha": Laughter as a Resource to Manage WhatsApp Conversations,' *Journal of Pragmatics*, 110: 1–19 <https://doi.org/10.1016/j.pragma.2017.01.001>

Pettersson, Rune. 2002. *Information Design: An Introduction* (Amsterdam: John Benjamins)

Pfalzgraf, Falco. 2009. *Englischer Sprachkontakt in Den Varietäten Des Deutschen - English in Contact with Varieties of German*, English in Contact with Varieties of German (Frankfurt am Main: Peter Lang)

Potts, Amanda. 2015. '"LOVE YOU GUYS (NO HOMO)" How Gamers and Fans Play with Sexuality, Gender, and Minecraft on YouTube,' *Critical Discourse Studies*, 12.2: 163–86 <https://doi.org/10.1080/17405904.2014.974635>

Poyatos, Fernando. 1993. *Paralanguage: A Linguistic and Interdisciplinary Approach to Interactive Speech and Sounds* (John Benjamins)

Preseau, Lindsay. 2018. '"We speak Görlienglisch": The Role of English in Multilingual/-ethnic German Youth Speech Communities,' in *Jugendsprachen: Aktuelle Perspektiven Internationaler Forschung*, ed. by Arne Ziegler (Berlin: De Gruyter), pp. 519–42

Pullen, Christopher, & Margaret Cooper. 2010. *LGBT Identity and Online New Media* (Abingdon: Routledge)

R Core Team. 2021. *R: A Language and Environment for Statistical Computing* (Vienna) <https://www.R-project.org/>

Ramachandran, Kandethody M., & Chris P. Tsokos. 2020. *Mathematical Statistics with Applications in R* (Academic Press)

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

Reershemius, Gertrud, & Evelyn Ziegler. 2015. 'Sprachkontaktinduzierte Jugendkulturelle Stile Im DaF-Unterricht: Beispiele Aus Dem Film „Fack Ju Göhte",' in *Interaktionale Sprache Und Und Ihre Didaktisierung Im*

*DaF-Unterricht*, ed. by Wolfgang Imo & Sandro M. Moraldo (Tübingen: Stauffenburg), pp. 241–74

Reif, Maria. 2012. *Verbzweitstellung Nach 'Weil': Ein Syntaktisches Phänomen Im Gesprochenen Deutsch* (Munich: GRIN Verlag)

Reis, Marga. 2013. '„Weil-V2"-Sätze Und (k)ein Ende? Anmerkungen Zur Analyse von Antomo & Steinbach (2010),' *Zeitschrift für Sprachwissenschaft*, 32.2: 221–62 <https://doi.org/10.1515>

Remus, Robert, Uwe Quasthoff, & Gerhard Heyer. 2010. 'SentiWS – a Publicly Available German-Language Resource for Sentiment Analysis,' in *Proceedings of the 7th Conference on International Language Resources and Evaluation (LREC'10)*, ed. by Nicoletta Calzolari et al. (presented at the LREC 2010, Valetta: European Language Resources Association (ELRA)), pp. 1168–71

Rinas, Karsten, & Veronika Uhrová. 2016. 'Perioden mit Smileys. Zum Verhältnis von Emoticons und Interpunktion,' *Linguistik Online*, 75.1 <https://doi.org/10.13092/lo.75.2519>

Riva, Giuseppe. 2002. 'The Sociocognitive Psychology of Computer-Mediated Communication: The Present and Future of Technology-Based Interactions,' *CyberPsychology & Behavior*, 5.6: 581–98 [accessed 2 March 2021]

Robertson, Alexander, Walid Magdy, & Sharon Goldwater. 2018. 'Self-Representation on Twitter Using Emoji Skin Color Modifiers,' in *Twelfth International AAAI Conference on Web and Social Media*

Rosenau, Heinz. 2001. 'Die Interaktionswirklichkeit des IRC,' *NetWorx*: 1–24

Rotne, Lena. 2018. '"I don't have time for tits." An Investigation of Italian and Danish Adolescents' Writing on Facebook and in School Essays,' in *Jugendsprachen: Aktuelle Perspektiven Internationaler Forschung*, ed. by Arne Ziegler (Berlin: De Gruyter), pp. 891–914

Ruesch, Jurgen, & Gregory Bateson. 1951. *Communication, the Social Matrix of Psychiatry.*, Communication, the Social Matrix of Psychiatry. (New York:

W. W. Norton & Company), p. 314

Runkehl, Jens, Peter Schlobinski, & Torsten Siever. 1998. 'Sprache Und Kommunikation Im Internet,' *Muttersprache*, 108: 97–109

Rymes, Betsy, & Andrea Leone-Pizzighella. 2018. 'YouTube-Based Accent Challenge Narratives: Web 2.0 as a Context for Studying the Social Value of Accent,' *International Journal of the Sociology of Language*, 2018.250: 137–63 <https://doi.org/10.1515/ijsl-2017-0058>

Røyneland, Unn. 2018. 'Virtually Norwegian,' in *Multilingual Youth Practices in Computer Mediated Communication*, ed. by Cecelia Cutler & Unn Røyneland (Cambridge: Cambridge University Press), pp. 145–67 <https://doi.org/10.1017/9781316135570.009>

Sacks, Harvey, & Gail Jefferson. 1992. *Lectures on Conversation* (Oxford: Blackwell)

Sacks, Harvey, Emanuel A. Schegloff, & Gail Jefferson. 1974. 'A Simplest Systematics for the Organization of Turn-Taking for Conversation,' *Language*, 50.4: 696–735

Saferinternet.at. 2018. 'Jugend-Internet-Monitor 2018,' *Saferinternet.at* <https://www.saferinternet.at/presse-detail/jugend-internet-monitor-2018/> [accessed 29 July 2018]

Sampietro, Agnese. 2016. 'Exploring the Punctuation Effect of Emoji in Spanish WhatsApp Chats,' *Lenguas Modernas*: 91–113

———. 2019. 'Emoji and Rapport Management in Spanish WhatsApp Chats,' *Journal of Pragmatics*, 143: 109–20 [accessed 10 February 2021]

Sassen, Claudia. 2000. 'Phatische Variebilität Bei Der Initiierung von Internet-Relay- Chat-Dialogen,' in *Soziales Im Netz: Sprache, Beziehungen Und Kommunikationskulturen Im Internet*, ed. by Caja Thimm (Wiesbaden: Westdeutscher Verlag), pp. 89–108

Savicki, Victor, Dawn Lingenfelter, & Merle Kelley. 1996. 'Gender Language Style and Group Composition in Internet Discussion Groups,' *Journal of*

*Computer-Mediated Communication*, 2 <https://doi.org/10.1111/j.1083-6101.1996.tb00191.x>

Saxalber, Annemarie, & Miriam Micheluzzi. 2018. 'Facebook-Sprachgebrauch Im Kontext von Innerer Mehrsprachigkeit in Südtirol,' in *Jugendliche Im Gespräch* (Bern: Peter Lang), pp. 277–98

Schäfer, Roland. 2015. 'Processing and Querying Large Web Corpora with the Cow14 Architecture,' in *Proceedings of the 3rd Workshop on Challenges in the Management of Large Corpora (CMLC-3)*. (Mannheim: Institut für Deutsche Sprache,), pp. 28–34

Schäfer, Roland, & Ulrike Sayatz. 2016. 'Punctuation and Syntactic Structure in Obwohl and Weil Clauses in Nonstandard Written German,' *Written Language & Literacy*, 19.2 (John Benjamins): 212–45

Scheffler, Tatjana. 2005. 'Syntax and Semantics of Causal Denn in German,' in *Fifteenth Amsterdam Colloquium* (Amsterdam: ILLC), pp. 215–20

Schegloff, Emanuel A., & Harvey Sacks. 1973. 'Opening up Closings,' *Semiotica*, 8.4: 289–327

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

———. 2001. '* Knuddel-Zurueckknuddel-Dich-Ganzdollknuddel*: Inflektive Und Inflektivkonstruktionen Im Deutschen,' *Zeitschrift für Germanistische Linguistik 29 (2001), Nr. 2*, 29.2: 192–218

———. 2005. 'Mündlichkeit/Schriftlichkeit in Den Neuen Medien,' in *Standardvariation: Wie Viel Variation Verträgt Die Deutsche Sprache?*, ed. by Ludwig Eichinger & Werner Kallmeyer (Berlin: De Gruyter), pp. 126–42

———. 2006a. 'Die Bedeutung Digitalisierter Kommunikation für Sprach- Und Kommunikationsgemeinschaften,' in *Von *Hdl* Bis *Cul8r*: Sprache Und Kommunikation in Den Neuen Medien*, ed. by Peter Schlobinski (Mannheim:

Dudenverlag), pp. 26–37

———. 2006b. *Von *Hdl* Bis *Cul8r*: Sprache Und Kommunikation in Den Neuen Medien* (Mannheim: Dudenverlag)

Schlobinski, Peter, Gaby Kohl, & Irmgard Ludewigt. 1993. *Jugendsprache: Fiktion Und Wirklichkeit* (Opladen: Westdeutscher Verlag)

Schmidt, Gurly. 2000. 'Chat - Eine Kommunikative Gattung,' in *Soziales Im Netz: Sprache, Beziehungen Und Kommunikationskulturen Im Internet*, ed. by Caja Thimm (Wiesbaden: Westdeutscher Verlag), pp. 109–30

Schmidt, Günter Dietrich. 1987. 'Das Affixoid. Zur Notwendigkeit Und Brauchbarkeit Eines Beliebten Zwischenbegriffs Der Wortbildung,' in *Deutsche Lehnwortbildung. Beiträge Zur Erforschung Der Wortbildung Mit Entlehnten WB-Einheiten Im Deutschen.*, ed. by Gabriele Hoppe et al. (Tübingen: Narr), pp. 53–101

Schneider, Edgar. 2013. 'Investigating Historical Variation and Change in Written Documents: New Perspectives,' in *The Handbook of Language Variation and Change*, 2nd edn, ed. by J. K. Chambers & Natalie Schilling-Estes (Oxford: Blackwell), pp. 57–81

Schwartz, H. Andrew et al. 2013. 'Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach,' *PLOS ONE*, 8.9 (Public Library of Science): e73791 <https://doi.org/10.1371/journal.pone.0073791>

Schwencke, Nena. 2012. 'Wissen sie, was sie tun? - Analyse des Sprachverhaltens und der Nutzungskompetenz auf SchülerVZ,' *NetWorx* <https://doi.org/10.15488/2953>

Schwitalla, Johannes. 2003. *Gesprochenes Deutsch: eine Einführung* (Berlin: E. Schmidt)

Scott, A. 2015. 'Sprachliche Variation im digitalen Zeitalter und ihre Relevanz für den DaF-Unterricht,' *Deutsch als Fremdsprache*, 3: 164–71

Scott, K. 2015. 'The Pragmatics of Hashtags: Inference and Conversational

Style on Twitter,' *Journal of Pragmatics*, 81: 8–20 [accessed 31 March 2021]

Seargeant, Philip. 2019. *The Emoji Revolution. How Technology Is Shaping the Future of Communication* (Cambridge: Cambridge University Press)

Searle, John R. 1976. 'A Classification of Illocutionary Acts,' *Language in Society*: 1–23

Sefton-Green, Julian. 1998. 'Introduction: Being Young in the Digital Age,' in *Digital Diversions: Youth Culture in the Age of Multimedia*, ed. by Julian Sefton-Green (London: Routledge), pp. 1–18

Shatz, Marilyn, Henry M. Wellman, & Sharon Silber. 1983. 'The Acquisition of Mental Verbs: A Systematic Investigation of the First Reference to Mental State,' *Cognition*, 14.3: 301–21 <https://doi.org/10.1016/0010-0277(83)90008-2>

Siebenhaar, Beat. 2006a. 'Code Choice and Code-switching in Swiss-German Internet Relay Chat Rooms,' *Journal of Sociolinguistics*, 10.4: 481–506

———. 2006b. 'Gibt Es Eine Jugendspezifische Varietätenwahl in Schweizer Chaträumen?' in *Perspektiven Der Jugendsprachforschung- Trends and Developments in Youth Language Research*, ed. by Christa Dürscheid & Jürgen Spitzmüller (Frankfurt am Main: Peter Lang), pp. 227–39

———. 2006c. '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

———. 2018. 'Funktionen von Emojis und Altersabhängigkeit ihres Gebrauchs in der WhatsApp-Kommunikation,' in *Jugendsprachen: Aktuelle Perspektiven Internationaler Forschung*, ed. by Arne Ziegler (Berlin: De Gruyter), pp. 749–72

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

Siever, Torsten, Peter Schlobinski, & Jens Runkehl (eds.). 2005. *Websprache.net: Sprache Und Kommunikation Im Internet* (Berlin: Walter de Gruyter)

Silverstein, Michael. 1976. 'Shifters, Linguistic Categories, and Cultural Description,' in *Meaning in Anthropology*, ed. by Keith Basso & Henry Selby (Albuquerque: University of New Mexico Press), pp. 11–55

———. 1993. 'Metapragmatic Discourse and Metapragmatic Function,' in *Reflexive Language: Reported Speech and Metapragmatics*, ed. by J.A. Lucy (Cambridge: Cambridge University Press), pp. 33–58

———. 2003. 'Indexical Order and the Dialectics of Sociolinguistic Life,' *Language & Communication*, 23.3-4: 193–229

Skovholt, Karianne, Anette Grønning, & Anne Kankaanranta. 2014. 'The Communicative Functions of Emoticons in Workplace E-Mails: :-)*,' *Journal of Computer-Mediated Communication*, 19.4: 780–97 <https://doi.org/10.1111/jcc4.12063>

Soffer, Oren. 2010. '"Silent Orality": Toward a Conceptualization of the Digital Oral Features in CMC and SMS Texts,' *Communication Theory*, 20.4: 387–404 <https://doi.org/10.1111/j.1468-2885.2010.01368.x>

Söll, Ludwig, & Franz Josef Hausmann. 1980. *Gesprochenes Und Geschriebenes Französisch* (E. Schmidt)

Spitzmüller, Jürgen. 2013. *Graphische Variation als soziale Praxis: Eine soziolinguistische Theorie skripturaler ›Sichtbarkeit‹* (Berlin: De Gruyter)

Spreckels, Janet. 2006. '"Tussis, Schlampen, Britneys Und Wir" Fremd- Und Selbstkategorisierung in Einer Adoleszenten Mädchengruppe,' in *Perspektiven Der Jugendsprachforschung- Trends and Developments in Youth Language Research*, ed. by Christa Dürscheid & Jürgen Spitzmüller (Frankfurt am Main: Peter Lang), pp. 151–64

Stark, Luke, & Kate Crawford. 2015. 'The Conservatism of Emoji: Work, Affect, and Communication,' *Social Media + Society*, 1.2 (SAGE Publications Ltd): 205630511560485 <https://doi.org/10.1177/2056305115604853>

Statista. 2020. 'UK: SMS and MMS Messages Sent 2019,' *Statista* <www.statista.com/statistics/271561> [accessed 2 June 2021]

———. 2021. 'Versendete SMS in Deutschland bis 2020,' *Statista* <www.statista.com/statistics/155052> [accessed 2 June 2021]

Stenson, Robert. 2020. '"TUNE IN/JOIN US": Mobilising Liveness as a Promotional Strategy in Film Trailer Exhibition' (Nottingham: University of Nottingham)

Stevenson, Patrick (ed.). 1995. *The German Language and the Real World: Sociolinguistic, Cultural, and Pragmatic Perspectives on Contemporary German* (Oxford: Clarendon Press)

Storrer, Angelika. 2001. 'Getippte Gespräche Oder Dialogische Texte? Zur Kommunikationstheoretischen Einordnung Der Chat-Kommunikation,' in *Sprache Im Alltag*, ed. by Lehr Andrea et al. (Berlin: De Gruyter), pp. 439–65

———. 2013. 'Sprachstil Und Sprachvariation in Sozialen Netzwerken,' in *Die Dynamik Sozialer Und Sprachlicher Netzwerke: Konzepte, Methoden Und Empirische Untersuchungen an Beispielen Des WWW*, ed. by Barbara Frank-Job, Alexander Mehler, & Tilmann Sutter (Wiesbaden: Springer), pp. 331–66

Strapagiel, Lauren. 2019. 'This Is Why VSCO Girls Keep Saying "Sksksksk

*Intelligent Text Processing*, ed. by Alexander Gelbukh (Berlin: Springer), 7817, pp. 121–36 <https://doi.org/10.1007/978-3-642-37256-8_11>

Tagg, Caroline. 2009. 'A Corpus Linguistics Study of SMS Text Messaging' (Birmingham: The University of Birmingham)

———. 2015. *Exploring Digital Communication: Language in Action* (Abingdon: Routledge) <https://doi.org/10.4324/9781315727165>

Tagliamonte, S. A., & D. Denis. 2008. 'LINGUISTIC RUIN? LOL! INSTANT MESSAGING AND TEEN LANGUAGE,' *American Speech*, 83.1: 3–34 <https://doi.org/10.1215/00031283-2008-001>

Tagliamonte, Sali A. 2016a. 'So Sick or so Cool? The Language of Youth on the Internet,' *Language in Society*, 45: 1–32

———. 2016b. *Teen Talk: The Language of Adolescents* (Cambridge: Cambridge University Press)

Tajfel, Henri, & John C. Turner. 1979. 'An Integrative Theory of Intergroup Conflict,' *The Social Psychology of Intergroup Relations*, 33: 74

Tannen, Deborah. 1994. 'Gender Gap in Cyberspace,' *Newsweek*, 123.20: 52–53

Teltarif. 2012. '1&1: All-Net-Flat & Internet - Alle Tarifdetails Auf Teltarif.de!' <https://www.teltarif.de/a/1und1/allnet-internet-jul11.html> [accessed 28 July 2021]

———. 2019. '1&1: All-Net-Flat LTE S - Alle Tarifdetails Auf Teltarif.de!' <https://www.teltarif.de/a/1und1/allnet-flat-lte-s.html> [accessed 28 July 2021]

ten Have, Paul. 2007. *Doing Conversation Analysis*, Introducing Qualitative Methods, 2nd edn (Los Angeles: SAGE)

Thaler, Peter. 2003. *Pictoplasma 2: contemporary character design* (Berlin: Gestalten Verlag)

The Unicode Consortium. 2011. 'Unicode 6.0.0'

<http://www.unicode.org/versions/Unicode6.0.0/> [accessed 14 April 2021]

———. 2021a. 'Blocks-14.0.0.txt' <https://www.unicode.org/Public/UCD/latest/ucd/Blocks.txt> [accessed 24 September 2021]

———. 2021b. 'Emoji List, V13.1' <https://www.unicode.org/emoji/charts/emoji-list.html> [accessed 9 March 2021]

———. 2021c. 'UAX #44: Unicode Character Database' <http://www.unicode.org/reports/tr44/#General_Category_Values> [accessed 24 September 2021]

Thompson, Dominic et al. 2016. 'Emotional Responses to Irony and Emoticons in Written Language: Evidence from EDA and Facial EMG,' *Psychophysiology*, 53 <https://doi.org/10.1111/psyp.12642>

Thompson, Dominic, & Ruth Filik. 2016. 'Sarcasm in Written Communication: Emoticons Are Efficient Markers of Intention,' *Journal of Computer-Mediated Communication*, 21.2: 105–20 <https://doi.org/10.1111/jcc4.12156>

Thorne, Steven L., & Dejan Ivković. 2015. 'Multilingual Eurovision Meets Plurilingual YouTube: Linguascaping Discursive Ontologies,' in *Dialogue Studies*, ed. by Dale A. Koike & Carl S. Blyth (Amsterdam: John Benjamins), pp. 167–92

Thurlow, Crispin, & Kristine R. Mroczek (eds.). 2011. *Digital Discourse: Language in the New Media* (Oxford: Oxford University Press)

Tolins, Jackson, & Patrawat Samermit. 2016. 'GIFs as Embodied Enactments in Text-Mediated Conversation,' *Research on Language and Social Interaction*, 49.2: 75–91 <https://doi.org/10.1080/08351813.2016.1164391>

Tovares, Alla V. 2019. 'Negotiating "Thick" Identities Through "Light" Practices: YouTube Metalinguistic Comments about Language in Ukraine,'

*Multilingua*, 38.4 (De Gruyter): 459–84

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

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

Uhmann, Susanne. 1998. 'Verbstellungsvariation in Weil-Sätzen: Lexikalische Differenzierung Mit Grammatischen Folgen,' *Zeitschrift für Sprachwissenschaft*, 17.1 (Walter de Gruyter, Berlin/New York): 92–139

Ulaby, Neda. 2006. 'OMG: IM Slang Is Invading Everyday English,' *NPR.org* <https://www.npr.org/templates/story/story.php?storyId=5221618> [accessed 6 April 2021]

Urhan Torun, Bahar. 2018. 'Z Kuşağının Akıllı Telefonlar Üzerinden Yazılı İletişimde Emoji Kullanma Eğilimlerine Yönelik Bir Araştırma': 18

Utz, Sonja. 2000. 'Social Information Processing in MUDs: The Development of Friendships in Virtual Worlds.' *Journal of Online Behavior* (Behavior Online)

Vandergriff, Ilona. 2013. 'Emotive Communication Online: A Contextual Analysis of Computer-Mediated Communication (CMC) Cues,' *Journal of Pragmatics*, 51: 1–12 <https://doi.org/10.1016/j.pragma.2013.02.008>

Variantengrammatik des Standarddeutschen. 2020. 'Wortstellung Im Verbalkomplex,' *Ein Online-Nachschlagewerk. Verfasst von Einem Autorenteam Unter Der Leitung von Christa Dürscheid, Stephan Elspaß Und Arne Ziegler* <http://mediawiki.ids-mannheim.de/VarGra/index.php/Wortstellung_im_Verbalkomplex> [accessed 4 September 2021]

Varis, Piia, & Tom van Nuenen. 2017. 'The Internet, Language, and Virtual Interactions,' in *The Oxford Handbook of Language and Society*, ed. by Ofelia García, Nelson Flores, & Massimiliano Spotti (Oxford: Oxford University Press), pp. 473–88

Vasiljevič, Anja. 2018. 'Jugendsprache und Facebook - Eine komparative

Untersuchung der deutschen und slowenischen Jugendsprache,' in *Jugendsprachen: Aktuelle Perspektiven Internationaler Forschung*, ed. by Arne Ziegler (Berlin: De Gruyter), pp. 915–28

Vellusig, Robert Heinz. 2000. *Schriftliche Gespräche: Briefkultur Im 18. Jahrhundert* (Vienna: Böhlau)

Vogelgesang, Waldemar. 2010. 'Digitale Medien – Jugendkulturen – Identität,' in *Digitale Jugendkulturen*, ed. by Kai-Uwe Hugger (Wiesbaden: VS Verlag für Sozialwissenschaften), pp. 37–53

Voigt, Martin. 2011. 'Soziolinguistische Studie Zur Beziehungsarbeit Und Identitätskonstruktion in Mädchenfreundschaften,' *NetWorx*, 61

———. 2015. *Mädchenfreundschaften unter dem Einfluss von Social Media Eine soziolinguistische Untersuchung* (Frankfurt am Main: Peter Lang) <https://www.peterlang.com/view/product/23163?format=EPUB> [accessed 27 July 2021]

Volodina, Anna. 2007. 'Konditionale Und Kausale Relationen Im Gesprochenen Deutsch' (Heidelberg: University of Heidelberg)

von Polenz, Peter. 1990. 'Nationale Varietäten Der Deutschen Sprache,' *International Journal of the Sociology of Language* (Walter de Gruyter, Berlin/New York Berlin, New York): 5–38

Waggoner, Erin B. 2018. 'Bury Your Gays and Social Media Fan Response: Television, LGBTQ Representation, and Communitarian Ethics,' *Journal of Homosexuality*, 65.13 (Taylor & Francis): 1877–91

Wagner, Franc, & Ulla Kleinberger. 2016. 'Reflexionen Zum Schreiben Jugendlicher in Neuen Medien,' in *Jugendsprache in Schule, Medien Und Alltag*, ed. by Carmen Spiegel & Daniel Gysin (Frankfurt am Main: Peter Lang), pp. 109–24 <https://www.peterlang.com/view/title/17633> [accessed 25 July 2021]

Walther, Diana. 2018. "'Doing Youth" – Zur Erweiterung einer Theorie der Jugendspracheforschung,' in *Jugendsprachen: Aktuelle Perspektiven*

*Internationaler Forschung*, ed. by Arne Ziegler (Berlin: De Gruyter), pp. 25–48

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 <https://doi.org/10.1111/j.1083-6101.2006.00305.x>

Watzlawick, Sonja. 2006. 'Jugendbiografien Und Sprachgebrauch Im Wandel Der Zeit,' in *Perspektiven Der Jugendsprachforschung- Trends and Developments in Youth Language Research*, ed. by Christa Dürscheid & Jürgen Spitzmüller (Frankfurt am Main: Peter Lang), pp. 281–94

Wegener, Heide. 1993. '"Weil - Das Hat Schon Seinen Grund": Zur Verbstellung in Kausalsätzen Mit "Weil" Im Gegenwärtigen Deutsch,' *Deutsche Sprache*: 289–305

———. 1999. 'Syntaxwandel Und Degrammatikalisierung Im Heutigen Deutsch?: Noch Einmal Zuweil-Verbzweit,' *Deutsche Sprache*: 511–44

Weida, Courtney Lee, & Carlee Bradbury. 2020. 'Magical Aesthetics of Unicorns in Girlhood Visual Culture,' *Visual Culture & Gender*, 15: 44–55

Weissman, Benjamin, & Darren Tanner. 2018. 'A Strong Wink Between Verbal and Emoji-Based Irony: How the Brain Processes Ironic Emojis During Language Comprehension,' *PLOS ONE*, 13.8 (Public Library of Science): e0201727 <https://doi.org/10.1371/journal.pone.0201727>

Werry, Christopher C. 1996. 'Linguistic and Interactional Features of Internet Relay Chat,' in *Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural Perspectives*, ed. by Susan C Herring (Philadelphia: John Benjamins), pp. 47–63

West, Candace, & Don H. Zimmerman. 1987. 'Doing Gender,' *Gender & Society*, 1.2: 125–51 <https://doi.org/10.1177/0891243287001002002>

Wharton, Tim. 2016. 'Paralanguage,' in *The Routledge Handbook of Pragmatics*,

ed. by Anne Barron, Yueguo Gu, & Gerard Steen (Oxford: Taylor and Francis), pp. 69–75 <https://research.brighton.ac.uk/en/publications/paralanguage> [accessed 8 February 2021]

WhatsApp. 2020. 'WhatsApp Help Center - How to Configure the Enter Key,' *WhatsApp.com* <https://faq.whatsapp.com/android/chats/how-to-configure-the-enter-key/?lang=en> [accessed 10 March 2021]

Wiese, Heike. 2009. 'Grammatical Innovation in Multiethnic Urban Europe: New Linguistic Practices Among Adolescents,' *Lingua*, 119: 782–806 <https://doi.org/http://dx.doi.org/10.1016/j.lingua.2008.11.002>

———. 2012. *Kiezdeutsch: Ein Neuer Dialekt Entsteht* (Munich: CH Beck)

———. 2015. '"This Migrants' Babble Is Not a German Dialect!": The Interaction of Standard Language Ideology and "Us"/'Them'dichotomies in the Public Discourse on a Multiethnolect,' *Language in Society*, 44.3: 341–68

———. 2022. 'Neue Dialekte im urbanen Europa,' in *Handbuch Sprache im urbanen Raum, Interdisziplinäre Perspektiven der Stadtforschung*, ed. by Busse Beatrix & Ingo H. Warnke (Berlin: De Gruyter) <https://www.degruyter.com/view/product/185122> [accessed 30 September 2019]

Wiese, Heike, & Ulrike Freywald. 2019. 'Regionalsprachliche Merkmale In Jugendsprachlichen Praktiken Im Multilingualen Urbanen Raum,' in *Deutsch: Sprache Und Raum - Ein Internationales Handbuch Der Sprachvariation*, ed. by Joachim Herrgen & Jürgen Erich Schmidt (Berlin: De Gruyter), pp. 995–1012 <https://doi.org/10.1515/9783110261295-037>

Wiese, Heike, Mehmet Tanir Öncü, & Philip Bracker. 2017. 'Verb-Dritt-Stellung Im türkisch-Deutschen Sprachkontakt: Informationsstrukturelle Linearisierungen Ein-Und Mehrsprachiger Sprecher/Innen,' *Deutsche Sprache*: 31–52

Wiese, Heike, & Ines Rehbein. 2016. 'Coherence in New Urban Dialects: A

Case Study,' *Lingua*, 172: 45–61

Willem, Cilia, Núria Araüna, & Iolanda Tortajada. 2019. 'Chonis and Pijas: Slut-Shaming and Double Standards in Online Performances Among Spanish Teens,' *Sexualities*, 22.4: 532–48 <https://doi.org/10.1177/1363460717748620>

Wilson, Deirdre, & Dan Sperber. 2002. 'Relevance Theory,' in *Handbook of Pragmatics*, ed. by Gregory Ward & Handbook of Pragmatics Horn (Oxford: Blackwell), pp. 606–32

Wirth, Uwe. 2005. 'Chatten. Plaudern Mit Anderen Mitteln,' in *Websprache.net: Sprache Und Kommunikation Im Internet*, ed. by Torsten Siever, Peter Schlobinski, & Jens Runkehl (Berlin: De Gruyter), pp. 67–84

Witmer, Diane F., & Sandra Lee Katzman. 1997. 'On-Line Smiles: Does Gender Make a Difference in the Use of Graphic Accents?' *Journal of Computer-Mediated Communication*, 2 <https://doi.org/10.1111/j.1083-6101.1997.tb00192.x>

Wittenberg, Eva, & Kerstin Paul. 2009. '"Aşkim, Baby, Schatz..." Anglizismen in Einer Multiethnischen Jugendsprache,' in *Englischer Sprachkontakt in Den Varietäten Des Deutschen = English in Contact with Varieties of German*, ed. by Falco Pfalzgraf (Frankfurt am Main: Peter Lang), pp. 95–122

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

Wolfer, Sascha, Carolin Müller-Spitzer, & Maria Ribeiro Silveira. 2019. '... Weil V2 Geht Auch: Empirische Untersuchungen Zu Angemessenheit Und Grammatikalität' (presented at the WiGBam: Was ist Grammatikalität?, Bamberg)

Wotanis, Lindsey, & Laurie McMillan. 2014. 'Performing Gender on YouTube: How Jenna Marbles Negotiates a Hostile Online Environment,' *Feminist Media Studies*, 14.6: 912–28

Wyss, Eva L., & Barbara Hug. 2016. 'WhatsApp-Chats. Neue Formen der

Turn-Koordinationbei räumlich-visueller Begrenzung,' in *Jugendsprache in Schule, Medien und Alltag*, ed. by Carmen Spiegel & Daniel Gysin (Frankfurt am Main: Peter Lang), pp. 259–74 <https://doi.org/10.3726/978-3-653-04950-3>

Yao, Mike Z, & Rich Ling. 2020. '"What Is Computer-Mediated Communication?"—An Introduction to the Special Issue,' *Journal of Computer-Mediated Communication*, 25.1: 4–8 <https://doi.org/10.1093/jcmc/zmz027>

Yates, Simeon J. 1996. 'Oral and Written Linguistic Aspects of Computer Conferencing,' in *Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural Perspectives*, ed. by Susan C Herring (Philadelphia: John Benjamins)

YouTube. 2020. 'Videos: Rate | YouTube Data API | Google Developers' <https://developers.google.com/youtube/v3/docs/videos/rate> [accessed 25 August 2021]

———. 2021. 'YouTube Creators | Awards,' *YouTube Creators | Awards* <https://www.youtube.com/creators/awards/> [accessed 13 August 2021]

*YouTube Premium Is Broken...* 2021. <https://www.youtube.com/watch?v=NLO5-IyS5pE> [accessed 25 August 2021]

Yus, Francisco. 2011. *Cyberpragmatics: Internet-Mediated Communication in Context* (Amsterdam: John Benjamins)

Yuval-Davis, Nira. 2011. *The Politics of Belonging: Intersectional Contestations* (London: SAGE)

Zannettou, Savvas et al. 2018. 'The Good, the Bad and the Bait: Detecting and Characterizing Clickbait on YouTube,' in *2018 IEEE Security and Privacy Workshops (SPW)*, pp. 63–69 <https://doi.org/10.1109/SPW.2018.00018>

Zappavigna, Michele. 2015. 'Searchable Talk: The Linguistic Functions of Hashtags,' *Social Semiotics*, 25.3 (Informa UK Limited): 274–91

<https://doi.org/10.1080/10350330.2014.996948>

Zifonun, Gisela. 1987. *Kommunikative Einheiten in der Grammatik*, Forschungsberichte des Instituts für Deutsche Sprache Mannheim, Bd. 65 (Tübingen: Narr)

———. 2000. 'Grammatische Integration Jugendsprachlicher Anglizismen,' *Der Deutschunterricht*: 69–79 <https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/4834> [accessed 20 July 2021]

# A

# Appendix: Statistical tables of the *NottDeuYTSch* corpus

Table A.1: List of channels included in the *NottDeuYTSch* corpus

| **YouTube channel names** | | | |
|---|---|---|---|
| 1LIVE | DieLochis | Lamiya Slimani | Sami Slimani |
| AGGRO.TV | Dilara Duman | LauraJoelle | Sarazar |
| AlexiBexi | Dima | LeFloid | Shirin David |
| Ana Lisa | E IL o T IR i X ™ | Lena | Simon Desue |
| Anna Maria Damm | Ebru Ergüner | Leon Machère | Sonny Loops |
| Anne Wünsche | Emrah | Lisa-Marie Schiffner | SophiaThiel |
| Anni The Duck | FabTheGap | MRS BELLA | TZON4life |
| ApeCrime | Felix von der Laden | Maren Wolf | TheBeauty2go |
| ApeCrimeTV | FloWest | MefYou | Tina Neumann |
| ApoRed | Freshtorge | Meggyxoxo | Tobias Wolf |
| Applewar | Hatice Schmidt | Melina Sophie | Tugay |
| AshMadeOurEyes | HeyMoritz | Michael Buchinger | ViktoriaSarina |
| BULLSHIT TV | Isabeau | Michael Schulte | World Wide Wohnzimmer |
| BUSHIDO | Ischtar Isik | Mirco Rosik | YTITTY |
| BarbaraSofie | JONAS | MixX - Die Klassiker | albertoson |
| BibisBeautyPalace | Jarow | Nihan | atzenmusiktv |
| Bonnytrash | Joyce | Nilam | bademeisterTV |
| Chameen | Julia Beautx | Niloofar Irani | coldmirror |
| Chimperator Channel | Julien Bam | Paluten | diejungs |
| ConCrafter \| LUCA | Julienco | Pamela Rf | flyinguwe |
| DASDING | KAYEFTV | Paola Maria | iBlali |
| DIE TOTEN HOSEN | KWiNK | PietSmiet | inscope21 |
| Dagi Bee | Katharina Damm | Pietro Lombardi | kitthey |
| DasMirkoo | KeysJore | Promiflash | madametamtam |
| Dennis Bro | KrappiWhatelse | RTL | manniac |
| Der Ömsen | KsFreakWhatElse | RayFox | ungespielt |
| Diana zur Löwen | LIONTTV | Rebekah Wing | xKarenina |
| DieAussenseiter | LIONTVLOG | Sam Masghati | xLaeta |

Table A.2: Number of videos selected for the *NottDeuYTSch* corpus per category and year

| Category/Year | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Autos & Vehicles | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | **3 (1.01%)** |
| Comedy | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | **11 (3.72%)** |
| Education | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 0 (0%) | 0 (0%) | **7 (2.36%)** |
| Entertainment | 1 (0.34%) | 1 (0.34%) | 2 (0.68%) | 4 (1.35%) | 4 (1.35%) | 4 (1.35%) | 6 (2.03%) | 7 (2.36%) | 12 (4.05%) | 24 (8.11%) | 14 (4.73%) | **79 (26.69%)** |
| Film & Animation | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | **10 (3.38%)** |
| Gaming | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 4 (1.35%) | 12 (4.05%) | 19 (6.42%) | 16 (5.41%) | 13 (4.39%) | 9 (3.04%) | 6 (2.03%) | 5 (1.69%) | **86 (29.05%)** |
| Howto & Style | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 2 (0.68%) | 3 (1.01%) | 3 (1.01%) | 3 (1.01%) | 3 (1.01%) | 4 (1.35%) | 2 (0.68%) | **24 (8.11%)** |
| Music | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | **11 (3.72%)** |
| News & Politics | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | **10 (3.38%)** |
| Nonprofits & Activism | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 0 (0%) | 0 (0%) | **4 (1.35%)** |
| People & Blogs | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 2 (0.68%) | 2 (0.68%) | 2 (0.68%) | 2 (0.68%) | 2 (0.68%) | 1 (0.34%) | **16 (5.41%)** |
| Pets & Animals | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 1 (0.34%) | 0 (0%) | 0 (0%) | **1 (0.34%)** |
| Science & Technology | 0 (0%) | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 0 (0%) | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 0 (0%) | **6 (2.03%)** |
| Shows | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 2 (0.68%) | 1 (0.34%) | 1 (0.34%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | **7 (2.36%)** |
| Sports | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | **11 (3.72%)** |
| Travel & Events | 0 (0%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | 1 (0.34%) | **10 (3.38%)** |
| **Total** | **6 (2.03%)** | **12 (4.05%)** | **14 (4.73%)** | **19 (6.42%)** | **28 (9.46%)** | **36 (12.16%)** | **37 (12.5%)** | **34 (11.49%)** | **37 (12.5%)** | **44 (14.86%)** | **29 (9.8%)** | **296 (100%)** |

Table A.3: Number of comments per video category and year in the *NottDeuYTSch* corpus

| category | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Autos & Vehicles | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 3 508 (0.11%) | 2 306 (0.07%) | 704 (0.02%) | **6 518 (0.21%)** |
| Comedy | 434 (0.01%) | 439 (0.01%) | 2 986 (0.09%) | 1 686 (0.05%) | 5 405 (0.17%) | 9 326 (0.3%) | 23 731 (0.75%) | 10 279 (0.33%) | 25 447 (0.81%) | 36 908 (1.17%) | 13 116 (0.42%) | **129 757 (4.12%)** |
| Education | 0 (0%) | 29 (0%) | 196 (0.01%) | 300 (0.01%) | 139 (0%) | 19 (0%) | 115 (0%) | 2 470 (0.08%) | 28 046 (0.89%) | 5 422 (0.17%) | 3 914 (0.12%) | **40 650 (1.29%)** |
| Entertainment | 494 (0.02%) | 2 590 (0.08%) | 7 236 (0.23%) | 13 854 (0.44%) | 22 588 (0.72%) | 70 378 (2.23%) | 39 747 (1.26%) | 81 248 (2.58%) | 340 062 (10.8%) | 348 523 (11.07%) | 203 871 (6.47%) | **1 130 591 (35.9%)** |
| Film & Animation | 0 (0%) | 86 (0%) | 3 781 (0.12%) | 3 204 (0.1%) | 3 503 (0.11%) | 1 056 (0.03%) | 4 561 (0.14%) | 2 124 (0.07%) | 4 373 (0.14%) | 4 044 (0.13%) | 2 069 (0.07%) | **28 801 (0.91%)** |
| Gaming | 0 (0%) | 24 (0%) | 118 (0%) | 883 (0.03%) | 49 482 (1.57%) | 89 822 (2.85%) | 78 498 (2.49%) | 97 392 (3.09%) | 124 231 (3.94%) | 70 262 (2.23%) | 30 754 (0.98%) | **541 466 (17.19%)** |
| Howto & Style | 5 (0%) | 380 (0.01%) | 3 585 (0.11%) | 5 119 (0.16%) | 5 188 (0.16%) | 15 998 (0.51%) | 52 969 (1.68%) | 63 920 (2.03%) | 323 636 (10.28%) | 345 654 (10.98%) | 51 115 (1.62%) | **867 569 (27.55%)** |
| Music | 0 (0%) | 93 (0%) | 1 219 (0.04%) | 849 (0.03%) | 1 610 (0.05%) | 1 723 (0.05%) | 2 227 (0.07%) | 1 210 (0.04%) | 5 446 (0.17%) | 35 710 (1.13%) | 7 436 (0.24%) | **57 523 (1.83%)** |
| News & Politics | 0 (0%) | 86 (0%) | 2 699 (0.09%) | 416 (0.01%) | 195 (0.01%) | 195 (0.01%) | 339 (0.01%) | 46 (0%) | 3 390 (0.11%) | 1 071 (0.03%) | 1 654 (0.05%) | **10 091 (0.32%)** |
| Nonprofits & Activism | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 184 (0.01%) | 18 (0%) | 68 (0%) | 1 571 (0.05%) | 142 (0%) | 34 (0%) | **2 017 (0.06%)** |
| People & Blogs | 25 (0%) | 818 (0.03%) | 566 (0.02%) | 3 405 (0.11%) | 6 290 (0.2%) | 12 501 (0.4%) | 7 258 (0.23%) | 16 762 (0.53%) | 52 608 (1.67%) | 77 513 (2.46%) | 16 871 (0.54%) | **194 617 (6.18%)** |
| Pets & Animals | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 1 053 (0.03%) | 8 (0%) | 3 (0%) | **1 064 (0.03%)** |
| Science & Technology | 0 (0%) | 0 (0%) | 32 (0%) | 54 (0%) | 20 (0%) | 6 (0%) | 12 (0%) | 51 (0%) | 986 (0.03%) | 1 101 (0.03%) | 11 (0%) | **2 273 (0.07%)** |
| Shows | 0 (0%) | 245 (0.01%) | 4 896 (0.16%) | 5 747 (0.18%) | 26 482 (0.84%) | 37 874 (1.2%) | 1 951 (0.06%) | 1 775 (0.06%) | 1 865 (0.06%) | 1 152 (0.04%) | 616 (0.02%) | **82 603 (2.62%)** |
| Sports | 11 (0%) | 56 (0%) | 170 (0.01%) | 565 (0.02%) | 948 (0.03%) | 1 520 (0.05%) | 1 801 (0.06%) | 479 (0.02%) | 5 801 (0.18%) | 11 819 (0.38%) | 2 368 (0.08%) | **25 538 (0.81%)** |
| Travel & Events | 0 (0%) | 26 (0%) | 138 (0%) | 198 (0.01%) | 204 (0.01%) | 384 (0.01%) | 2 079 (0.07%) | 4 956 (0.16%) | 12 149 (0.39%) | 6 272 (0.2%) | 1 973 (0.06%) | **28 379 (0.9%)** |
| **Total** | **969 (0.03%)** | **4 872 (0.15%)** | **27 622 (0.88%)** | **36 280 (1.15%)** | **122 054 (3.88%)** | **240 986 (7.65%)** | **215 306 (6.84%)** | **282 780 (8.98%)** | **934 172 (29.66%)** | **947 907 (30.1%)** | **336 509 (10.68%)** | **3 149 457 (100%)** |

Table A.4: Statistical overview of the *NottDeuYTSch* corpus

| **Statistic** | **Value** |
|---|---|
| Number of Tokens (including emoji and emoticons) | 33 760 494 |
| Number of Tokens (only lexemes) | 32 549 462 |
| Number of Types | 567 086 |
| Type-Token Ratio (TTR) | 0.017 |
| Number of Comments | 3 149 457 |
| Number of Videos | 296 |
| YouTube Channels Represented | 63 |
| Mean Tokens per Comment | 10.72 |
| Median Tokens per Comment | 5 |
| Mean Comments per Video | 1 914 |