“Dieses Unterlases ist zu einem Kreiswichs geworden”

An SFL-based investigation of the Anti-language of r/kreiswichs

“Dieses Unterlases ist zu einem Kreiswichs geworden”

Eine SFL-basierte Untersuchung der Anti-Sprache von r/kreiswichs

by

Jeffrey N. Lapalme

A thesis

presented to the University of Waterloo

and the Universität Mannheim

in fulfillment of the

thesis requirement for the degree of

Master of Arts

in

Intercultural German Studies

Waterloo, Ontario, Canada / Mannheim, Germany, 2016

Authors Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

Abstract

Reddit is a social-news-aggregator and online community with a large and active user-base from around the world. Although the main language of discourse on the site is English, Reddit's highly diverse membership has enabled and fostered the growth of numerous language communities, each with their own specialised subreddits, group dynamics, and linguistic norms. One of the oldest and largest of these communities is the German-language community centred around the r/de subreddit. In recent years, a relatively small but dynamic group, has sprung up out of the German-language community to create their own subreddit r/kreiswichs, one which is 'free' from Anglicisms, and where any and every term can and should be translated into German. Using both quantitative and qualitative means within a framework of Systemic Functional Linguistics, I analyse a corpus of kreiswichs user submission and comment data from the calender year 2015 with the goal of investigating whether their unique take on the German language and internet terminology meets the criteria of an anti-language as initially set out in Halliday (1976). To that end, I compare the lexicon of kreiswichs to that of the broader German-language Reddit community, analyse the qualitative elements of it, before moving on to its implementation and the discourse of user submissions in the subreddit.

Acknowledgements

First and foremost I would like to extend my deepest thanks and gratitude to Dr. Mathias Schulze for his guidance throughout this long process. Without his knowledge, insight, enthusiasm, and, most importantly, patience, I could not have made my way through this project. I would also like to thank Jason Baumgartner, first for taking it upon himself to create the Reddit Public Data Archive, and second, for taking the time to show me how to make sense of it all. Next, I must thank the untold masses of latex and linux experts on stackexchange.com and across the web, who by sharing their knowledge with the world, shared it with me, and without whom I would still be trying to compile files. I also extend my thanks to my readers for approaching my work with an open and critical mind. To my dear friend, Brannon Kelly, thank you for all your editing and thoughtful comments. To my family, friends, and fiancée, thank you all for keeping me around and putting up with me through this whole thing.

Dedication

Den Wichsern. Ohne euch und eure feuchte Maimais wäre das ganze hier nicht möglich. Hochwähls bitte zur Linken.

Contents

List of Figures

List of Tables

1 An Introduction to this Thesis: Why Kreiswichs?

For those familiar with Reddit, or a knowledge of the German language, the subject of this thesis may seem peculiar; what is the academic value in analysing one relatively small sub-community of one of thousands of different online discussion boards, forums, social networks, and chatrooms. Over the course of this work, I hope to make the scholarly value of my pursuit clear, or at a minimum convince you of what I already feel to be the truth of the matter, which is that this community is incredibly interesting, due in no small part to their lexical creativity. The group at the centre of this work engage in a discourse which is at times absurd and fanciful, and at others bitingly satirical and insightful. The members of r/kreiswichs (KW) demonstrate an amazing degree of linguistic competence, enabling them to manipulate languages to make complex, nuanced meanings. These meanings which may not require any particularly high level of metalinguistic awareness to understand, but certainly in order to appreciate them.

My interest in studying KW is linked to my long-standing interest in anonymous online discourse-communities. Over several years of casual observation, and indeed participation, I grew to wonder how exactly it is that these totally unknown groups of individuals develop a sense of community and connection to one another? This connection seems to form even though none of those involved can truly know to whom they are communicating with or whether or not replies are indeed coming from the same users. I have also wondered how this anonymity impacts the languages used in the discourse itself. Given the nature of truly anonymous online discussion boards such as 4chan.org, it is exceedingly difficult to develop and analyse detailed timelines

and corpora of these interactions. KW, on the other hand, as a pseudo-anonymous community with automatically archived posts, eliminates some of these major difficulties. Here we can see individuals grow to know each other by 'name', interact and develop discourses over the course of months and years. We can see rules develop and how they are applied. In short, KW provides an example of linguistic phenomena in a realm between open identity and total anonymity. This can then be used to further our understanding of those extreme ends of the same scale.

Tied to my desire to learn more about the interplay between anonymity and language, I feel there is a need for the recording of the online culture of memes (see chapter 2). These snippets of culture reflect the shifting moods, opinions and subcultures which have been spawned by, and within, what is undoubtedly the greatest technological wonder of the 20^{th} century. The global communities of the internet are, in the grand scheme of things, still very much in their infancy, and these early years are something that we ought not to let go undocumented. This is our chance to document the life cycles of entire cultures. I hope that through this work I will be able to contribute my part to this task.

Taking a step back from more grandiose notions, to those which are perhaps more practical, KW is a German-language community revolving largely around the written word supplemented by other modalities. As such it provides an excellent opportunity to examine linguistic theories developed with other forms of mediums and channels of communication in mind (e.g. spoken, telephonic, etc). We can also test theories which have to date largely only been examined in the context of the English language, allowing us to identify their potential strengths or weaknesses as generalised linguistic theories.

Ultimately, KW represents a subject which offers an astounding array of

potential avenues of investigation for researchers and scholars of a wide range of disciplines - even if the name may be a bit off-putting to some.

1.1 The Research Question and Objective of this Thesis

The research question guiding this thesis is a fairly straight-forward one - Is the language of the kreiswichs subreddit an anti-language? Despite its apparent simplicity, in order to be able to answer this question, several steps must be followed. It is these steps which also make up the objectives of this work. First, we must establish that if there is indeed a difference in the language used by the KW-community to that of the larger German-language Reddit community. Second, if this difference does exist, we must seek to understand the nature of it, i.e. what exactly makes this language unique. Third, once we have uncovered the qualities of this language, we must move on and determine how is it used and how KW members react to it versus more 'standard' or differing varieties of the language. Finally, we must determine if the if these characteristics of the language, its reception, and use are consistent with the concept of anti-language.

In the following pages and chapters, sections and subsections, I describe my efforts to answer this question in a manner consistent with good academic practice. Furthermore, I hope to do justice to the subject of my thesis, that is, the members of KW and the creativity with which they manipulate languages, bending German and English to meet their needs. While the discourse in which they engage is largely very humorous, filled with equal parts pointed satire, wit, and absurdity, it is also one worthy of a thorough and thoughtful analysis.

1.2 Theoretical Foundations

This thesis and the analyses therein have been built around the framework provided by Systemic Functional Linguistics (SFL). This is a theory of language pioneered by M.A.K. Halliday beginning in the early 1960s, and has since become a highly influential and comprehensive theory of language use. I will delve more deeply into the theory in later chapters, in particular chapter 3, but the major tenets of SFL can be (very) briefly described as follows: *Language is a system of choices, and understanding these choices gives us insight into how language functions as a tool for making meaning in our social environment.*

Further underpinning my work is the aforementioned concept of an anti-language, also developed by Halliday. Despite what may be suggested by the name, anti-language is in this usage not something (or someone) *against* language, but rather it is a language itself. An anti-language exists in an oppositional relationship to the language from which it stems (see chapter 3, section 3.1).

Given SFL's origin as a theory developed by same researcher who would later conceptualise anti-language, theory and concept are well positioned to complement each other. This is, of course, to say nothing of SFL's robustness as an 'appliable' (see chapter 3, section 3.2) theory of language, a theory which has been continuously developed through the decades, and shown its value as an effective tool for understanding the sociolinguistic world around us.

1.3 Ethical Considerations & a Note about Terminology

Given the nature of this research, which looks at the language use of a specific group of individuals, I would like to discuss the ethics surrounding the data collection and analysis thereof. This thesis makes use of linguistic data collected as part of an archive of public data (see chapter 4) without the knowledge or explicit consent of the participants. These data consist of nearly all[1] *publicly* available data posted on Reddit by its users since the creation of the site; there are so-called "private" subreddits, but these are not available to the general public, nor are they included in these data. Collected in the archive along with the content of a post is the user name, including any *flair* of the user, the time and date it was posted, the name of the subreddit on which the content was posted, and the *karma* and *gilded* score of said post.[2] These data, when analysed together, may enable one to discover personally identifiable information about a given user. That said, *all* of this information is freely available not only to the registered user-base of Reddit, but the general public as well, as even non-registered visitors can access it. This is covered in Reddit's privacy policy which is readily available to the public (Reddit, 2016b) and users of the site are encouraged to familiarise themselves with its details prior to posting content.

As stated above, individual usernames are available within the datasets, i.e. the corpora developed from this archived information, however they are the creation of the user themselves and do not reflect their actual identities unless they have chosen for this to be the case. Again, the public nature of usernames is explicitly outlined in the Reddit privacy policy (Reddit, 2016b).

[1] Some 300 thousand posts could not be collected due to various technical issues.
[2] For descriptions of these terms, please refer to chapter 2, section 2.1.

Although this thesis makes use of an independently compiled archive, any individual with sufficient coding knowledge or access to purpose-made programs can collect and archive these data, which has been made available through Reddit's own Application Programming Interface (API). Reddit provides access to its API freely, and hosts documentation for those interested in making use of it (Reddit, 2016a).

Given the explicitly public nature and availability of these data, the research undertaken in this course of thesis is, according to Chapter 2, Articles 2.2 and 2.3 of the Research Ethics Board of Canada's Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (Pre.ethics.gc.ca, 2014) and the University of Waterloo's own ethics policies (Research, 2014), exempt from the ethics review process, and does not violate any expectations of privacy on behalf of the subjects.

I would also like to mention that over the course of this thesis, several terms, usernames and phrases are introduced which may be offensive to some readers. I have taken care to avoid any appearance of gratuitous use of potentially offensive terminology via the use of acronyms and initialisms where possible. That said, in order to remain true to the subject matter, my sources, and to provide an honest accounting of the phenomena at play, the correct terminology must be used. This is also true of any examples of discourse which are used for illustrative purposes in this thesis.

1.4 Structure of this Thesis

This thesis has been divided into six chapters. The first chapter is a brief overview of all information pertinent to understanding the goals and flow of this thesis, including the motivation, research question and theoretical foundation upon which it has been built, as well as a brief discussion of the ethics

of this study. Chapter two introduces the subject of this study and outlines several key concepts which, while not tied to the theories employed in this work, are nonetheless critical to understanding the subject and subject matter moving forward. Chapter three is the a review of relevant literature as it relates to anti-language and SFL. In chapter four, the datasets used in this thesis are described in detail as are the methods used to analyse them. The fifth chapter is a presentation and discussion of the results of my analyses, and with them, an answer to the research question. The final chapter concludes this thesis with a summary of the key findings along with some of the limitations of this work. It also highlights opportunities for further study, both with regards to the subject matter and the data used.

1.5 Conventions, Annotations and Definitions

Over the course of this work, particularly in chapters 4, and 5, you will encounter system network diagrams (see chapter 3) and annotated text. These diagrams follow the typical conventions of SFL which are as follows:

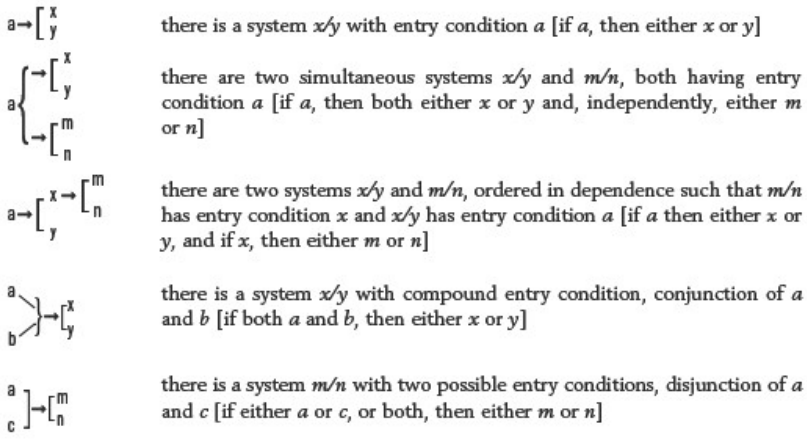


Figure 1.1: System mapping conventions adapted from Halliday & Matthiessen (2013).

Systems are typically mapped in a horizontal fashion, starting at the left and expanding to the right as the network grows. Arrows are used to indicate the direction of expansion as well. Each node or point along the system is considered an expansion of the *delicacy* of the system, as each feature highlights a nuance of the feature preceding it:

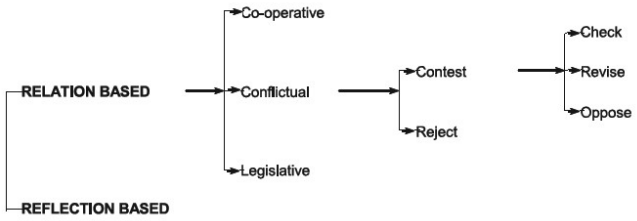


Figure 1.2: A map of features within a subsystem. Adapted from Wegener (2011).

When presented as written text, rather than as a system network, major variables and their features have been written in SMALL CAPS. The symbol → is used to indicate that successive features represents a further level of delicacy from the feature preceding it:

RELATION BASED → CONFLICTUAL → CONTEST → OPPOSE

Figure 1.3: A written example of hypothetical selections from within the subsystem in figure 1.2.

For features which occur simultaneously at the same level of delicacy + is used:

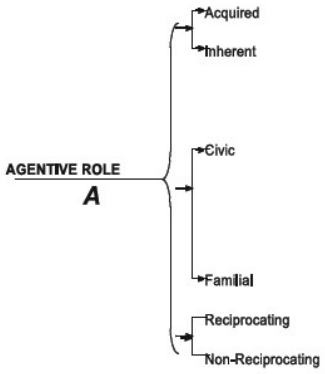


Figure 1.4: A map of features within a subsystem exhibiting simultaneous options. Adapted from Wegener (2011).

AGENTIVE ROLE → AQUIRED + CIVIC + NON-RECIPROCATING

Figure 1.5: A written example of hypothetical selections from within the subsystem featured in figure 1.4.

Within SFL, there is a specific conceptualisation of 'context' which differs in definition from the perhaps more colloquial usage of the term. The word 'context' occurs throughout this thesis, and in order to highlight those instances when what is meant is the SFL concept, 'Context' will be used. In instances when 'context' is to be understood according to its more common usage, it will appear as either as 'situational context' or written solely in lower-case letters. For a description and discussion of 'context' as defined within SFL, please see chapter 3, section 3.2.

In this thesis, the term 'discourse' is used frequently. This is a word which has seen widespread application in the literature of numerous fields and disciplines, and for that reason there are a number of possible definitions one may encounter. In this thesis, discourse is defined as text(s) generated dynamically

(Fawcett, 2009) by individuals to make and convey related meanings. These meanings may be related due to their overarching themes or topics, their situational context, or the connections, i.e. relationships, between participants.

Now that I have given a brief outline the subject, motivations, foundations, and the structure of this work, let us move on to chapter two. There I will begin the more detailed discussions of topics and themes relevant to my thesis, and as described above, introduce some key concepts.

2 Important Concepts and Background Information

Before diving into academic and theoretical aspects of this thesis, there are several key concepts which are necessary for us to understand the environment in which the subject of this thesis is situated. These concepts, namely the "circle-jerk," and "shit posting," can take on a great many forms, in terms of not only their ostensible vulgarity, as is the case with the more widely familiar concept of the *meme* but also their content, themes, and reception. Unless otherwise specified, the term meme, used here and throughout the thesis, refers to the internet meme. These phenomena can be viewed as a subset of the concept of memes, first introduced by Dawkins (1976). Dawkins describes memes as 'units' of culture which are transmitted, imitated, and replicated amongst, and by, individuals and groups (Dawkins, 1976). As such, it can be difficult to create over-arcing descriptions of these phenomena. In fact it may be easiest to paraphrase Justice Potter Stewart (1997) - *you will know it when you see it.*[1] However, I will endeavour to provide as concise a description as possible, in order to provide the needed background for this thesis.

2.1 Reddit: The Frontpage of the Internet

Founded on June 23rd, 2005, by Alexis Ohanian and Steve Huffman (Wikipedia, 2016b), Reddit has since become one of the internet's most popular websites,

[1]This quote stems from the case *Jacobellis v. Ohio* in which the defendant, Nico Jacobellis, appealed his conviction for the crime of obscenity due to his exhibition of the film *The Lovers (Les Amants)*. Jacobellis' conviction was ultimately overturned, and in Justice Stewart's concurrence with the decision he stated "I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description; and perhaps I could never succeed in intelligibly doing so. But I know it when I see it, and the motion picture involved in this case is not that."

ranked 9th in the U.S. and 33rd globally[2] according to the online traffic ranking organisation Alexa (Alexa, 2016). In terms of raw numbers, Reddit boasts some 231 Million unique visitors per month[3] from 210 countries around the world (Reddit, 2015b).[4] At any given time there can be as many as 2.8 Million users logged into the site, though this number changes constantly with the time of day and day of the week.

The website itself is what is known as a social-news-aggregator (SNA). The general concept of an SNA is to provide its user-base with a place to post content, both their own[5] and content found elsewhere on the internet, as well as discuss posts and comments submitted by other users of the site. Reddit is by no means the first or only SNA, with other notable examples being Digg.com and Fark.com. As with other SNAs the content on Reddit is organised according to topics and themes.

Reddit does not have a homepage *per se*, but is made up of subreddits. which are then aggregated to what is called the frontpage. This frontpage is known as "/r/all", and provides an aggregated view of subreddits selected by Reddit staff. This is the default webpage for all visitors of the site, and is displayed to users prior to their becoming a registered member and creating their own custom frontpage. Subreddits are pages dedicated to specific topics, and run the gamut from politics, to specific countries or cities, themes, celebrities, tv shows, adult content, and nearly anything else which might interest an individual. With the exception of several major subreddits, for example *funny*, *AskReddit*, and *pics*, all subreddits are user created and moderated. Should a particular interest not be represented or a user be dissatisfied with an exist-

[2]Current as of 21.03.2016.

[3]This number is updated monthly and is based on the previous month's data. This is current as of 21.03.2016.

[4]No information is provided as to how this statistic is obtained. Presumably, it is based on user IP addresses, which indicate the origin of the connection.

[5]user created content is known as OC - *Original Content*.

ing subreddit, they can create a new subreddit, decide its content rules, and attempt to generate a user base. Within the discourse amongst Reddit users, who are referred to as *redditors*, specific subreddits are generally mentioned alongside */r/* (or often simply *r/*. as will be the case in this thesis), which represents a portion Reddit's URL structure which indicates that what follows is a subreddit. Thus, the subreddit *funny* would be referred to as */r/funny*, *AskReddit* is */r/AskReddit*, and so on.

There are a number of acronyms which users of Reddit will encounter while browsing the site, and which might not be familiar to new or casual users of the internet. The most common acronyms are laid out in the Reddit wiki along with a detailed compilation of frequently asked questions (FAQ) regarding to site policy, administration, user and moderator privileges and/or duties, subreddit creation, and so on (Reddit, 2015a). It should be noted that the acronyms listed in the FAQ represent only an incomplete list, and that they are in no way unique to Reddit, but are commonly encountered on other websites, forums, and in other forms of digital communication. The same applies to other concepts described in the following subsections. However, for the sake of consistency, examples will be framed within Reddit.

Now that Reddit has been introduced, I will now provide some more details about the demographics of its user-base, before moving on to a description of the subjects of my analyses, and finally some important, non-Reddit-specific concepts.

The Demographics

Reddit is a primarily Anglophone website and community. Approximately 64.02% of all web traffic to the site originating from Australia (2.3%), Canada (4.2%), Ireland (0.5%), the United Kingdom (5.5%), and the United States

(51.7%) according to Alexa Internet inc. (Alexa, 2016). Reddit ranks within the top 20 most popular websites in all of the aforementioned countries;[6] the only primarily non-anglophone countries where that is also the case are Sweden and Singapore, where Reddit ranks 16th, and 17th, respectively.[7]

Truly reliable and detailed demographics for Reddit as a whole are difficult to come by. However, according to statistics generated by Alexa's Global Traffic Panel, one can see certain demographic tendencies of the site compared to the company's assessment of global internet demographics. Based on their calculations, males are slightly over-represented within the Reddit userbase, as are individuals ages 18 to 34, individuals without children, individuals with no or some post-secondary education,[8] and individuals earning $60,000 or more per year. Individuals of Caucasian ethnicity are highly over-represented on Reddit, while all other ethnicities[9] are under-represented to various degrees, with individuals of Middle-Eastern descent constituting the least represented group of users. Although Alexa is a well established analytics company, the lack of transparency regarding their data collection methods and the statistics generated from them limit the confidence that one should place in said data for scientific purposes. They have been presented here, however, as they represent one of the few credible sources for said information. Data from Pew Research Center studies conducted in 2013 (Duggan & Smith, 2013) and 2016 (Barthel et al., 2016) show consistent findings, albeit purely in the context of American users.

[6]The popularity rankings are as follows: United States 9, Canada 13, Australia 15, Ireland 17, United Kingdom 17.

[7]Together, only 1.6% of all internet traffic on Reddit are generated by these two countries.

[8]In the terminology used by Alexa - 'No College' and 'Some College'.

[9]How ethnicities are determined or classified by Alexa is not made available; "African" and "African American" recorded separately.

Karma, Gilding & Post Ranking

One of the most identifiable and central features of Reddit is its voting system. When a user posts a comment or a submission within a subreddit, the content can be voted on by other users. The votes are known as either "upvotes" or "downvotes" and the sum of the upvotes minus the downvotes are referred to as the post's "karma." Karma is cumulative for users but is separated into "link karma" and "comment karma", and can be considered to reflect a user's standing within the overall (sub)Reddit community. Karma scores also determine the visibility of posts; comments and submissions with a high net karma score are filtered to the top of threads for maximum exposure, while low karma posts being relegated further down and may require users to directly click on a comment to read it. Comments which users feel are worthy of particular (positive) recognition can be "gilded." Gilding is when a user awards another user with what is called "Reddit gold." The post which is gilded receives a small icon of a golden coin next to the user's name along with an indication of how many times that post has been gilded, should it be more than once. Reddit gold is purchased with actual money and awards gilded users with certain benefits, like the removal of ads from the user interface, and increased website customisation (Reddit, 2016c). Certain types of posts, "self posts" can accumulate up- and downvotes, with the corresponding influence on post visibility, but they do not count towards a user's karma scores. Comments and submissions can be sorted by users based on a number of criteria, for example "hot", which is current post popularity,[10] "new," "rising," which is a ranking of post according to their increasing popularity, "controversial," which arranges posts based on the relative balance between

[10]This is the default sorting structure for content posted to Reddit, both within subreddits and individual threads. How the ranking algorithm works has been changed over time. For details regarding the current algorithm see Salihefendic (2015).

the number of up- and downvotes, "top," which ranks post according to their total karma accumulation within a user-defined timespan, and "gilded," which sorts comments according to their gilded status.

It should be noted that karma scores may vary slightly as a user refreshes any given submission page. According to the Reddit wiki FAQ page "the vote numbers are not "real" numbers, they have been "fuzzed" to prevent spam bots etc. So taking the above example, if five users upvoted the submission, and three users downvote it, the upvote/downvote numbers may say 23 upvotes and 21 downvotes, or 12 upvotes, and 10 downvotes. The points score is correct, but the vote totals are 'fuzzed"' (Reddit, 2015a). This is because such manipulation could unduly influence the course of discussions or be used to artificially flood the most prominent sections of the site with content. For example, a company could program a bot to upvote all content related to their products, while downvoting their competitors. If the numbers are "fuzzed", to use Reddit's terminology, a voting bot such as this would be unable to determine if their votes are being registered which may cause them to attempt to vote again which would nullify their first vote if it was indeed registered, as activating the same voting button twice undoes the vote, potentially sending the bot into a kind of voting loop. This would in turn aid Reddit administrators in blocking bots because their systems would notice the repeated votes from a single 'user' on a specific submission.

2.2 Subreddits

As mentioned in section 2.1, Reddit is made up of subreddits, which are dedicated to wide range of broad themes and specific topics. In the following section I will provide an overview of the subreddits used a point of reference against which the data from KW was compared, as well as the reasons for

their selection.

r/de

The subreddit was first established on February 20, 2006, and is among the oldest subreddits on the site. Only the subreddits r/olypmics, r/features, r/request, r/nsfw, r/sub and r/reddit.com are older (steve, 2006; Redditmetrics, 2016a), although the general discussion Japanese and Russian language subreddits, r/ja and r/ru, respectively, were established the same day. The name 'de' refers to the primary language of the board and hints at the reason for its creation, which was to provide a space for German to be the primary language of discourse, and with it, for the discussion of issues which are of particular interest to German speakers. Although German is the primary language of the discourse on r/de, English speakers are also welcome to participate.

There are currently 31,893 users[11] subscribed to r/de, and it is ranked as the 1500th largest subreddit according to redditmetrics.com.[12]

The discourse on r/de is multifaceted, with a wide breadth of topics and themes being discussed at any given point. In an effort to help users navigate the subreddit, and find themes which interest them, discussions are tagged with a number of different labels[13], several of which can be filtered into or out of the main page of the subreddit. The sidebar of de also provides links to many other subreddits which are either primarily German-language in focus or deal with topics related to German-speaking regions, as well as a German-language IRC[14] chat group.

[11] Current as of 24.04.2015

[12] The site is a tool for tracking subreddit creation and growth, and is not affiliated with Reddit inc. or its parent company

[13] Categories include: Wissen & Technik, Essen & Trinken, Politik, Verschwrung, Diskussion, NewsAT, NewsCH, NewsDE, and several others.

[14] IRC stands for Internet Relay Chat, and is protocol for text-based internet communication.

Due to r/de's broad range of primarily German-language discourse, as well as its high level of user activity, it was selected as the primary reference subreddit for the comparative aspects of my analysis. Indeed, r/de far exceeds the size of all subreddits included in my analysis, including the main subject of this thesis, KW (see section 4.3).

r/SCHLAND

r/SCHLAND is a circlejerk (CJ) style (see subsection 2.4) subreddit, with a focus on "everything that is awesome about Germany"(freshoutofthebox, 2013). The discourse on r/SCHLAND is primarily humorous in nature. Common themes are positive stereotypes of Germans and Germany, lists on which Germany ranks highly, satires and parodies of trending topics, and mainstream German culture. As is common of CJ-subreddits, there are memes which are unique to r/SCHLAND. For example Germany is referred to as *SCHLAND*, and adjective *schländisch* is used in place of *deutsch*. This is reference to the subreddit r/MURICA and to the slang term/stereotype which provided its name.[15] Also typical of CJ-subreddits, the upvote icon has been replaced, in this case with a German flag. While submissions can only be upvoted, comments can be downvoted as well; the downvote icon has been replaced with an image of the Danish flag.

The language used on r/SCHLAND is primarily German, though English is also welcome. The only explicit rule in regards to content is that no Nazi-related posts are permitted.

Established on February 23, 2013, r/SCHLAND is a smaller and relatively new subreddit with 6,106 subscribers. This places it as the 5,991st most

[15]The stereotype referenced here is that of the overly-patriotic american, whose southern accent which alters, and in the case of the initial /a/ removes, the pronunciation of the vowels in the word 'America'.

popular subreddit (Redditmetrics, 2016c).

r/SCHLAND is the largest, primarily German-language CJ-subreddit and serves to round out the reference corpus (see chapter 4) with discourse which is thematically CJ in nature.

r/self_de

The subreddit, r/self_de, is a comparatively tiny subreddit with only 444 subscribed members. The subreddit is geared towards providing the same experience as 'r/self', whose only limiting factor is that there can be no linking to other content outside of Reddit,[16] and users are expected to be courteous. Unlike r/self, r/self_de, only allows submissions and comments in German or dialects thereof.

r/self_de was first created on February 19, 2013, and has grown slowly, if steadily, since then. It currently ranks 29,455th in terms of subscribers.

Despite its small size r/self_de was selected due to the broad range of possible themes and topics of discussion. It is also the only solely German-language subreddit, with no posts in any other language.

2.3 r/kreiswichs

The kreiswichs subreddit was created on November 20, 2012 by the redditor GuantanaMo. This user also submitted the first post entitled *auch wenn euch das, nicht passt, ich wähle bei den nächsten bundestagswahlen die piratenpartei!*, which was a self post, linking to no outside content or providing any further information or elaboration. The post received three comments

[16]This is a feature of the so-called 'self post.' These posts can be made on any subreddit, but r/self and r/self_de are explicitly and exclusively for this format.

and approximately[17] 14 upvotes. Over the following two years after this post, KW would grow slowly, generally gaining one or two subscribers per day, and occasionally losing them as well. The first major increase in terms of members came on May 10, 2014 when the number of subscribers rose by 257% to 625. This marked the beginning of more rapid growth for KW, which saw two further relatively large increases on January 27, 2015 (268 additional subscribers) and March 7, 2015 (127 additional subscribers). Currently, there are 3,625 subscribers[18].

The name and general format of KW is borrowed from the English-language subreddit, /r/circlejerk,[19] which was founded nearly four years earlier on December 10, 2008. KW is an overwhelmingly humour oriented subreddit. The vast majority of submissions and comments are parodies and satires directed at current events, on all levels, from global, to the comments within individual submissions on KW itself. The r/de subreddit (see section 2.2) is a particularly common target for the members of KW, and it is not uncommon for them to *raid* the subreddit, SPing (see section 2.4), and generally disrupting its members as they try to engage in their normal discourses. Beyond directing humour towards others, the members of KW are also self-deprecating, referring to each other as *Wichsers*[20] and mocking both themselves and their fellow *Mitwichser*. Members of KW are also active outside of reddit, and have formed "semi-official"[21] gaming groups and IRC-chat groups; an analysis of these groups is, however, outside the scope of this thesis.

With regards to the individual members of KW, it is difficult to put together a detailed image of who exactly they are. This is, of course, due largely

[17] For an explanation as karma scores can only generally given as approximates please refer to subsection 2.1.

[18] Current as of 03.05.2016.

[19] For an explanation of the name please see: section 2.4.

[20] English equivalents would be terms along the lines of *jerk-off* or *wanker*.

[21] Translation of the author.

to the semi-anonymous nature of Reddit's membership system; user's may give themselves whatever names they wish, and can use and create new accounts on a whim. This leads to obvious difficulties in determining accurate demographics. One can only assume that the members of KW are part of the German-speaking subset of Reddit's typical member-base (see above). So what can in fact be said of the *Wichsern*? For one, there is a relatively small group of user's which are highly active on the subreddit. Of the over 3,300 members of KW, only 1478[22] users posted 1 or more comments or submissions. Furthermore, only 25 of those users posted more than 200 times, compared to some 581 users with only one post. Looking more closely at the top posters, a clear core of users emerges, amongst whom Wumselito is by far the most active. This user account alone is responsible for nearly 10% (1991) of all submissions and comments on r/kreiswichs over the course of 2015. The next most active user is Pitikay (1035 posts), with GuantanaMo (693), quillsandsofas (672), LeSpatula (656), Klarname (653), BombastixderTeutone (618), Admiral_STAHLGENITAL (564), manaburn777 (485), and EndOfTheDigitalAge (444). Beyond these data relating to the frequency by which a particular account posts, little more can be discussed in great detail. What is clear, as will be shown in later chapters, these individuals possess an extensive knowledge of both German and Anglophone (pop) culture, as well as a high degree of linguistic competence; it goes beyond the scope of this work to analyse the content posted by individual users across Reddit in order to create a detailed demographic account than I have provided here. These data are, however, available both within the RPDA and elsewhere for those wishing to conduct further research.

Turning our attention back towards the KW-community more generally,

[22] An indeterminate number of users are only listed as [deleted] in the KW-corpora, as such they can only be accounted for as a single user, bringing the total number to 1479.

we can see that it, as is a defining element of CJ subreddits, KW creates and fosters a number of memes that are unique to their subreddit, and which the members refer to as *maimais* (see chapter 5, section 5.5). Of particular interest to members are Seilbahnen, comedian Jan Böhmermann, and Aiman Abdallah, the host of the science-oriented television show Galileo. All three feature heavily on KW, in particular Aiman, who is referred to as *Maimann*, a play-on-words between *maimai* and *Aiman*. Aiman's image appears consistently in submissions and comments, and his face has been digitally manipulated and combined with a cartoon image of the sun which is used as the custom background for the subreddit. Böhmermann's face can also be seen floating in a cloud in the same image.[23]

The feature of KW which is of particular interest and a significant focus of this thesis is the extensive creation and use of loan-translations derived from English-language terminology of not only Reddit, but the greater internet as a whole. Dozens of lexical items have been translated, in some cases directly as calques (see chapter 5, section 5.2), and in others, by applying new meanings to existing items. In many instances, these new lexical items are also created to replace items already integrated into the German language, and in some cases have been codified by KW users in a *sticky post* known as the *Begrüßungsfaden*[24] which provides some translations for those new users who are unfamiliar with the terms. This particular submission, now somewhat out-of-date, was also a location where users could suggest alternative translations to the common (English) Reddit terminology which was otherwise codified in the body text of the post.

[23] Current as of 03.05.2016.

[24] A permanent post, message, or thread, on an online messaging board. *Stickies*, as they called, are usually kept at the top level of a given discussion board so as to draw user attention to them, and contain information which all users are expected to know.

2.4 Important Concepts not Limited to Reddit

Circlejerk

The concept of a "circlejerk"[25] within the situational context of online communication and interaction is central to this thesis. There is a clear sexual origin and association with this term, which is often implicitly referred to when used outside of that situational context. As a label it is generally applied pejoratively to describe the discourse of online communities, and refers to the constant re-cycling and repetition of memes, jokes, and the practice of "karma whoring"[26] (Don, 2013). The phenomena can be likened to that of the "echo chamber", whereby the discourse of a group is broadly characterised by others as a self-reinforcing feedback loop. This leads to discourse which can be difficult for the uninitiated to participate in; jokes and topics may appear opaque with regards to their humour, focus, or relevance.

While this term has negative connotations, it has been embraced by many communities within Reddit and the internet more generally. A great deal of subreddits have appropriated the characterisation, and have been founded either explicitly under the circlejerk (CJ) moniker[27] or as a place to allow this kind of discourse. Topics ranging from specific media forms and formats, political ideologies, arts and sciences, to geographical locations (i.e. continents, countries, and cities), and even "meta" discussions of Reddit itself are represented by their own CJ subreddits. Humour, irreverence, parody, satire, and often biting critiques of accepted norms are staples of CJ discourse, particu-

[25] Also often spelled: "circle-jerk", or "circle jerk".

[26] Karma whoring refers to users submitting content which panders "to the to the stereotypical prejudices or trends that are widely accepted by its members, including activities like reposting popular content and linking to websites with overwhelmingly positive reputation" (B, 2012).

[27] Many subreddits append either *circlejerk* or *jerk* to their topic of interest as a way of explicitly identifying the style of discourse to be found in said subreddit: r/gamingcirclejerk; r/beardjerk.

larly when they 'target' ideologies and how they are represented within the discourse of the greater Reddit or online community.

Shitposting

A phenomenon commonly encountered in comment threads across internet platforms and communities, *shitposting* (SPing) and *shitposts* (SP) refer to what are considered to be low quality or low effort posts by users. SPs are often made intentionally as a form of trolling, whereby they are used to derail and disrupt threads (Mercer & Zed, 2014). SPs can take a wide variety of forms, and can range from the use of specific words and phrases, images, and ASCII[28] characters (Mercer & Zed, 2014).

A large portion of what moderators are tasked with can be considered to be the prevention and removal of SPs, and the punishment of accounts associated with the practice. However, not all SPing is met with scorn by redditors (or moderators), and is permitted should it conform to the format of a particular subreddit. Indeed, there are many subreddits, where SPing is considered to be central element of the discourse found there; CJ-subreddits in particular are seen as a place where SPing, if not always explicitly sanctioned, is implicitly condoned and promoted by users. One could reasonably argue that CJ subreddits are a means of providing a safe space for this behaviour, as well as a means of containment of the same, which in turn provides a kind of protection for the larger discourse surrounding a particular topic.

Since SPing is such a central aspect of CJ discourse, it can be hard to categorise or highlight it as such. This is because if a discourse community promotes SPing, be it explicit or implicitly, posts of that nature lose a part of what make them disruptive and therefore, in sense, what makes them a SP.

[28] ASCII stands for American Standard Code for Information Interchange and is one of several standards for encoding characters for use and display by computing systems.

This is not to say that a 'true' SP is not possible in the situational context of a CJ, and in fact there is a very real level of meta-awareness among some users as to what makes a post a SP, and posts which are entirely fitting in CJ discourse are often viewed as SPs because they would be considered such by the greater discourse community.

In an attempt to create a standard framework for categorising posts as SPs, I have highlighted several characteristics which I believe point to a posting as being such. A post need not have all of these characteristics, and situational context is very important; however, the more of these criteria which a post fits, the more certain you can be that it is in fact a SP:

1. Low-effort - The post is simplistic in terms of its content i.e. limited to a few words, ACSII characters, or image macros. - Use of the 'User-Simulator'[29] and other bots fall into this category due to the minimal effort required to invoke them, and the quality of the posts which they produce.

2. Obstructive - It is designed to consume an inordinate amount of screen space.

3. Copypasta - A "copypasta" is a block of text which is recycled and can be used in a number of situations. The name comes from the act of "copying and pasting" text. Copypastas range from the widely known to the relatively obscure and can be 'created' from any given text. Both the use of copypasta, and the individual copypastas themselves are memes in their own right.

[29]The UserSimulator is a 'bot' program created by a redditor who goes by the username Trambelus. Users can call up the bot which then creates a generic post using text mined from the comment history of a designated user.

4. Canned-phrase(s) - Similar to copypastas, but shorter, these phrases contribute little to nothing in the way of advancing dialogue and are often simply meme-phrases in and of themselves.

5. Repetitive - Repetition of the same combination of characters, be they text or image.

6. Irrelevance - The post is contextually out of place in relation to the overall topic of the thread or subreddit.

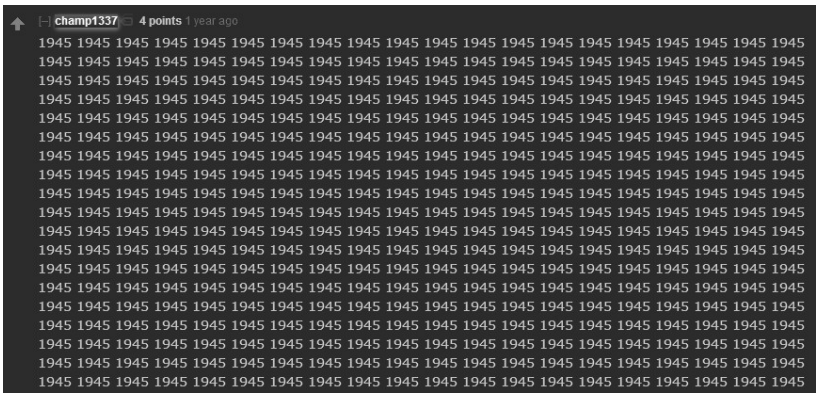


Figure 2.1: An excerpt from an SP by user Champ1337. Several characteristics are exhibited including: irrelevance, obstruction, and repetition.

2.5 Reddit Enhancement Suite

The final element of background information that I would like to introduce is Reddit Enhancement Suite (RES). RES is an unofficial, open-source third-party extension available for most modern web-browsers, and was developed primarily by Steve Sobel. The purpose of RES is to increase the functionality available to Reddit members when browsing the website. Among many helpful functions relating to account management and content filtering, it also enables users to scroll through subreddit submissions as if they were posted on a single,

long webpage, rather than requiring users to load pages separately. This function does, however, keep track of which page a submission is posted on according to the typical filter functions of Reddit (e.g. top, new, controversial). This particular feature of RES is not only convenient as a user of the site but also plays a role in this thesis (see chapter 4).

Now that we have covered the background information required to understand the framework surrounding the subject of this thesis and the larger community in which it is situated, we can move on to the theories and concepts which form the academic framework of this work. In the following chapter I will introduce the concept of anti-language, and Systemic Functional Linguistics, and provide of a review of the literature on the same.

3 A Review of the Literature

3.1 Halliday's Concept of Anti-Language

In his 1976 paper 'Anti-Languages', published in the journal *American Anthropologist*, M.A.K. Halliday introduced his conception of the languages used by societies which have arisen as a concious alternative to the larger societies within which they find themselves (Halliday, 1976). He termed these societies *anti-societies*, and their languages, *anti-languages*. These concepts are complementary to one-another, just as society and language are complementary; the pair that is (anti-)language and (anti-)society is "an instance of the prevailing sociolinguistic order" (Halliday, 1976, p. 570) - the two are inseparable from, and reinforcing of, one another.

Anti-Society

Anti-societies are largely composed of individuals or groups which are marginalised, or perceive themselves to exist on the fringes of a larger society due any number of possible reasons. This may be due to their lifestyles which are viewed by others to be undesirable or damaging (see Halliday, 1976; Bolton & Hutton, 1995; Villoria, 2011), or because of ideological differences (Champagne, 1977; Al-Azmeh, 1991). However, this is not to imply that the distancing of members of an anti-society from the larger society is always a process of externally imposed exclusion, as is the case with some criminal anti-societies (see Podgrecki, 1973; Bolton & Hutton, 1995). Often the separation between society and anti-society is self-imposed by members of an anti-society seeking to preserve what they feel makes their group distinct, or because they themselves view the larger society to be in some way corrupted (Schniedewind, 1999).

Anti-societies arise as a form of resistance towards a prevailing society. The resistance of the anti-society may be realised in a number of forms ranging from a kind of "passive symbiosis" to outright hostility and destructive action aimed towards the given society (Halliday, 1976, p. 570). Regardless of the form(s) of resistance expressed by an anti-society, it is, in essence, dependant upon the prevailing society for its existence; an anti-society is an extension of, and simultaneously a metaphor for, the society which it resists (Halliday, 1976, p. 578). The metaphoric nature of anti-societies can be seen in their alternate social structures. These structures "appear as a distorted reflection of the structure of the particular society from which it derives" (Halliday, 1976, p. 573), with its own values and systems of punishment and reward, and acceptance and gratification (Halliday, 1976, p. 575).

A number of communities have been classified as anti-societies in various literature; Halliday himself provides three examples which he feels typify the concept. In his paper, he highlights works describing the 'criminal classes' of Elizabethan England, the subcultures of prisons and reform schools in Poland, and the 'Calcutta underworld' as presenting clear examples of anti-societies (as well as anti-languages, which they make extensive use of) (Halliday, 1976, 570). As in Halliday's paper, a large portion of the groups discussed in anti-society literature have been those connected in some way to criminal groups (see Rediker, 1981; Villoria, 2011; Bolton & Hutton, 1995; Beier, 2005), or criminality more generally (see Chesney, 1970). However the phenomenon is not limited to criminal communities, by any means. The community of early Christians were considered by Halliday (Halliday, 1976, p. 575) to be that of an anti-society, a conceptualisation which has seen some acceptance amongst theologians who have adopted it while conducting their own investigations of early Christian communities (see Groenewald, 2011). Scholars studying

the Dead Sea Scrolls and the ancient Jewish community to whom they are attributed have also applied the framework to a limited extent as well (see Domeris, 1999).

Anti-Language

Anti-language is a phenomenon intrinsically tied to that of an anti-society, as an anti-language exists as a means of resocialisation and the realisation of alternate, subjective realities for said anti-society (Halliday, 1976, pp. 575-576). The basis for his assertion that anti-languages generate alternate realities is found in Berger (1966) which highlights the critical role of conversation and the "conversational apparatus" as a tool for the construction, maintenance, and modification of an individual's subjective reality (Berger, 1966, pp. 172-173). Anti-language's existence, and the reality which it creates and maintains is one of opposition, whereby the norms of the anti-language are established in contrast to existing norms of the language-reality that it resists. Thus, the social meanings inherent within any given anti-language are evaluated in oppositional terms; values are defined not by what they are, but rather what they are *not* (Halliday, 1976, p. 576). A further role served by the oppositional resocialisation of anti-language is the creation of strong affective ties and feelings of identification amongst its speakers - that is to say, amongst the members of an anti-society who in turn generate an anti-language (Halliday, 1976, p. 575). In the context of membership identification, there is generally a heavy reliance "on the foregrounding of interpersonal meanings, especially where . . . the cornerstone of the new reality is a new social structure- although, by the same token, the interpersonal elements in the exchange of meanings are likely to be fairly ritualized" (Halliday, 1976, p. 575). It should also be noted that due to the resocialising role of anti-language, it cannot function or

exist as a 'mother tongue'; anti-language is always essentially a reaction to a given language precluding it from achieving the kind of primacy typical of an L1.

Interestingly, it is "not the *distance* between the two realities but the *tension* between them which is significant" (Halliday, 1976, p. 576)(emphasis in the original). This is because the social structure foregrounded by the anti-language is, while different from the normative society which it resists, very much a part of the same social system; "they are variants of one and the same underlying semiotic" (Halliday, 1976, p. 576), and as such, there exists a continuity between language and anti-language. This continuity exists on the lexicogrammatical level of anti-language as well. In fact, a key aspect of the linguistic features of an anti-language is the relexicalisation, and in many cases an overlexicalisation, of socially relevant areas of a given language's vocabulary. The general grammar, however, remains unchanged (Halliday, 1976, p. 571) as do the areas of the lexicon which are irrelevant for the differentiation between the anti-society's members and the dominant society as a whole. This re- and over-lexicalisation is made clear when looking at the example of the anti-language of the Calcutta underworld as presented by Halliday. He notes some 21 different lexical items with the meaning "bomb", 24 synonyms for "girl", as well as some 41 items for the word "police" (Halliday, 1976, p. 571). Anti-languages, are, of course, not simply relexicalised variants of a given language; new vocabulary is introduced as well. For example, in records of the 'cant' spoken by the so-called criminal classes of Elizabethan England, the are some 20 new items which were introduced to denote the roles and positions of members of their society, and many more for the criminal acts themselves (Halliday, 1976, p. 571). As reasons for this relexicalisation and

introduction of vocabulary, Halliday suggests the need for secrecy[1] as well as demonstrations of verbal prowess as being part of the equation (Halliday, 1976, p. 572), along with the reality creation and maintenance discussed above.

Just as anti-societies are metaphors of the societies which they oppose, "anti-language is a metaphor for an everyday language" (Halliday, 1976, p. 578), which Halliday considers to be its defining characteristic. Halliday asserts that metaphor can be found at every level of the system, including at phonological, morphological, lexical, and semantic levels (Halliday, 1976, p. 578). In fact, while metaphor is by no means a feature exclusive to anti-language, it is the norm relative to the frequency of their occurrence in 'standard' languages. "Metaphorical compounding, metatheses, rhyming alternations, and the like" (Halliday, 1976, p. 579) are the expected features of the realisation of an anti-langauge.

Anti-language may be viewed in terms of a 'social dialect', however it does not fully encompass the scope of the phenomenon; Halliday notes parallels between the concepts of dialect and anti-language, while reaffirming its role as a language stating:

> The distinction between standard and non-standard dialects is one of language versus anti-language, although taking a relatively benign and moderate form. Popular usage opposes dialect, as "anti-," to (standard) language, as the established norm. A non-standard dialect that is consciously used for strategic purposes, defensively to maintain a particular social reality or offensively for resistance and protest, lies further in the direction of an anti-language... The social function of dialect variation is to express, symbolize, and

[1] For Halliday, secrecy is more a feature of the jargon, rather than anti-language as a whole.

> maintain the social order; and the social order is an essentially hierarchic one. An anti-language is, at one and the same time, *both* the limiting case of a social dialect (and hence is the realization of one component in the hierarchy of a wider social order that includes both society and anti-society), and a language (and hence the realization of a social order that is constituted by the anti-society itself); in the latter role, it embodies its own hierarchy, and so displays internal variation of a systematic kind (Halliday, 1976, pp. 580-581).

Halliday also argues for anti-language as text by highlighting that 'functional interpretation' of anti-language's components relating to the selection of meaning, the linguistic system, and "the situation, as a semiotic construct derivable from the 'social semiotic"'[2] (Halliday, 1976, p. 581). Unfortunately, Halliday does not elaborate on this further.

Applications of the Concept

'Anti-societies' are mentioned in several articles and other literature pre-dating Halliday's paper outlining his conceptualisation of the phenomena. That said, the concepts presented in those instances share a largely superficial similarity with the Hallidayan concept. They are either not fleshed out enough for one to draw a fair comparison to Halliday's work for example (cf. Lehman, 1968; Certeau & Brammer, 1992; Sprott, 1969), or are contextually quite different (cf. McNerney, 1969). While the defining features of anti-language and anti-society have not seen much in the way of further, explicit,[3] development, they

[2]The term social semiotic refers to the conceptualisation of culture as an system of meanings and meaning-making.

[3]Merrim (1980) makes mention of anti-language and a corresponding "anti-world" in her analysis of the language of the novel *Tres Tristes Tigres* but the idea is not elaborated on further in the course of her paper. Therefore, one can only speculate as whether or not this

have been applied to a number of fields of study beyond 'strict' sociolinguistics. These can be seen as contextual extension and development of the concepts, and evidence of its wide-ranging applicability.

Researchers and scholars of the humanities, broadly speaking, seem to have been particularly receptive to Halliday's concepts. Numerous papers published in various fields of literary studies have made use of, or reference to, either anti-societies, anti-languages, or both in the discussion of their respective subjects, ranging from analyses of texts such as in Kohn (2008), to the language of authors as represented in and by their writings. Stewart (1980), for example, examines the relationship between tradition and allusion in literary depictions of pickpockets and pickpocketing. In her paper, she compares anti-language to parody and satire, and concludes that these acts create their own societies (Merrim, 1980, pp. 1138-1139). The writings of author Gertrude Stein have been compared to anti-language in Stimpson (1985), whereby over the course of her personal development with regards to her own sexuality"Stein's coding of sexual activities ceases to be a suspect evasion and becomes, instead, a privileged, and a distinguished, 'anti-language"' (Stimpson, 1985, pp. 75-76).[4]

Anti-language in particular has found some currency amongst anthropologists and ethnologists. Reeves (1995) analyses the phenomenon of Muslim saint veneration and highlights the telling of *karama* (miracle) stories, the texts of which "may be understood as an 'anti-language' that reconstitutes with negative examples the image of interpersonal justice"(Reeves, 1995, p. 312). The language associated with the Zār rituals, which are ritual exorcism of individuals believed to be possessed by malevolent spirits practised in north-east Africa, has also been discussed within the framework of anti-

represents a true extension of Halliday's concept, or perhaps a metaphor for the alternate 'world', i.e. reality, of an anti-society.

[4]Stimpson's view is further discussed in De Lauretis (1988), which is an analysis of lesbian identity and "sexual difference" (De Lauretis, 1988, p. 155) as portrayed in literature.

language (Boddy, 1989). The Zār language enables the possessed, typically women, to express otherwise subversive feelings and ideas without fear of these outbursts being associated with them during their day-to-day lives within their community (Harris, 1997). Boddy's work influenced other investigations and interpretations of similar possession phenomenon in regions and times as diverse as the French border lands in the 19th century (Harris, 1997), and modern-day Kenya (Mcintosh, 2004). Both studies found parallels with their subject matter with that of Boddy's, however Harris, in particular finds that the "social commentary" function of anti-language to be insufficient to explain the possession phenomenon in the context of her study (Harris, 1997, p. 466). The concept has even been applied to fashion in Heath's (1992) analysis of Baay Fal clothing and style of dress found in Senegal, as well as to the discourse surrounding architectural ideology and theory in Oxman (2010).

There have also been numerous instances of the term 'anti-society' being used and described in Hallidayan terminology but without specific reference to Halliday himself. In his 1991 paper entitled *Islamist Revivalism and Western Ideologies*, Aziz Al-Azmeh describes actions of Sayyid Qutb, who advocated for "direct action" against supporters of western and secular ideologies,[5] the preparation for which "involve[d] a detachment from society and the formation of an anti-society by a spiritual and intellectual elite which works for radical subversion" (Al-Azmeh, 1991, p. 45) The term has also been applied to groups which took part in the Parisian protests of May, 1968, whose movement formed a "festive" anti-society which was "opposed to existing social relations and propos[ed] new ones" (Champagne, 1977). As no direct mention of Halliday's paper or related literature is made, it would be a tenuous proposition to assert that these interpretations are connected to his theory. However, one

[5]In particular the Egyptian government under Gamal Nassar.

can see that his concepts are rational enough that others would come to similar conclusions in their analysis of intra-societal dynamics.

3.2 Systemic Functional Linguistics (SFL) & Discourse Analysis

SFL was (and continues to be) developed as a comprehensive linguistic theory "concerned with language[6] in its entirety, so that whatever is said about one aspect also is to be understood always with reference to the total picture. At the same time, of course, what is being said of any one aspect also *contributes to* the total picture" (p.20 Halliday & Matthiessen, 2013, emphasis in the original). This is not to suggest that studies undertaken with SFL as the theoretical point of origin must provide a description of a given language as a whole, but rather that the features under investigation must be understood in context if one is to reach significant conclusions. SFL has its earliest roots in the work of anthropologist Bronislaw Malinowski, whose "insights were taken up and developed within linguistic theory by J.R. Firth" (Halliday & Matthiessen, 2013, p. 32). Firth's theories would then be further developed into what is now SFL, most notably by M.A.K. Halliday (e.g. 1967; 1978). While Halliday's influence on the development of SFL cannot be overstated, there are numerous researchers who have also contributed a great deal, among whom (socio)linguists such as Hasan (e.g. 1977; 1996; 2014), Martin (e.g.1991; 1995), and Matthiessen (e.g. 1988; 2012) have been arguably the most prolific.

At its core, SFL seeks to provide a framework for describing language in "particular, comparative, and typological" terms and is meant in this way to

[6]Halliday makes a point of noting that 'language' in the context of SFL is natural human, adult, language rather than the "proto-language of infants," semiotic systems such as computer and programming languages, mathematics, "languages of art" Halliday & Matthiessen (2013).

be an "appliable" theory (Webster, 2009, p. 9), which enables researchers to investigate the *semiotic*, or meaning-making, potential of language. The meaning-making potential is what simultaneously defines our experiences and enables the building and maintenance of social relationships. In keeping with this assertion, the conceptualisation of language as a 'social semiotic' is a central tenet of SFL (Halliday & Webster, 2009; Halliday & Matthiessen, 2013; Hasan, 2013). The influence of language on society, and society on language is reciprocal in nature, as language allows for social practices to be performed, while "the systemic aspects of culture ... legitimise the choice of meanings that can be meant appropriately in the context of that social practice" (Hasan, 2013, p. 274).

Just as culture is systemic, language is systemic in nature. Each feature of any given language represents an option or alternative for making meaning, which in turn determines the set of options (entry conditions) which are available afterwards. It is important to note that the term system is one with differing meanings in SFL depending on the situational context in which it is used. On the one hand it refers "to the system of language as a whole" (Hasan, 2013, p. 280), and, on the other, to the network of choices available when using language to make and exchange meaning. More generally the term system refers to the organising concept for modelling features within SFL, whereby every node in the system represents the point at which the possible selections must be made in order for the features of successive levels to be accessible.

Due to the systemic nature of language and the paradigm of features of which it consists, the notion of choice[7] is an intrinsic element of language. It is the choice of a particular feature or another of language which guides or

[7]The conceptualisation of choice and choosing within SFL has been the source of significant discussion. For a collection of detailed overviews and analyses of the issues surrounding this aspect of SFL see Fontaine et al. (2013).

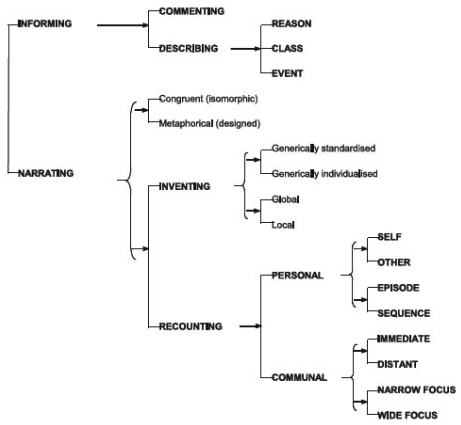


Figure 3.1: A system focused on features of narration. Adapted from Wegener (2011).

determines the paths which are taken as an individual navigates the system in order to make meaning. This is not to imply that a language user is consciously or intentional choosing each successive feature within the system as they make their way through it towards any given result. However, choice is "a convenient way of conceptualising the paradigmatic relationship between meanings that the language makes available to us" (Fawcett, 2013, p. 121).

While SFL frames language as a system of features and choices, it also emphasises that this system is stratified. The number and scope of these strata have been developed continuously throughout the ongoing evolution of SFL[8] (Matthiessen, 2009). Currently the typological model of language in SFL spreads it across four strata, whereby each successive, deeper level constitutes the means of realisation of the level preceding it. The first two levels are semantics and lexicogrammar, respectively. These strata represent the 'content' plane of language, and are what allows for language's meaning potential to grow "more or less indefinitely" (Halliday & Matthiessen, 2013, p.

[8]SFL is considered to be "an open dynamic system" (Matthiessen, 2009, p. 12) which is updated to keep pace with the phenomena with which it is concerned.

25). The 'expression' plane consists of the phonological and phonetic strata, and are the means by which humans can "interfac[e] with the environment" (Halliday & Matthiessen, 2013, p. 25). All of the above mentioned strata are encapsulated within the (social) context of a given language, which is to say the meanings as determined by the semiotic system of culture (Fontaine, 2013; Halliday, 2009; Hasan, 2013). The strata are distinct in that they represent different types of abstraction within language, but are equal in importance when attempting to describe the system of any given language as a whole (Hasan, 2013). The unifying force which binds the strata is that of the aforementioned 'realisation'. To put it another way, realisation is the hierarchy of abstraction through which meaning is exchanged across strata and, ultimately between individuals (Hasan, 2013; J. Martin, 2009). The term realisation also applies to the notion of choice in SFL (see above) whereby structural features, such as the selection of mood type,[9] interrogative type,[10] etc. are realisations of systemic choices, and expressions of the relationships between all aspects of language (Halliday & Webster, 2009).

Given the explicitly comprehensive nature of SFL, it is beyond the scope of this thesis to address every feature or function of language as conceptualised within the theory.[11] In the following subsections I will provide a overview of the SFL conceptions most relevant to my analysis, namely the interpersonal, ideational, and textual metafunctions, and the related concepts of FIELD, TENOR, and MODE (FTM). In SFL the term metafunction is used to describe groupings of related *semantic systems*, that is, systems which are related to the making and expression of meanings. The contextual variables

[9] For example, indicative, hypothetical or imperative.

[10] for example, yes/no or 'wh-' interrogative.

[11] For those interested in more detailed accounts of SFL as a whole, see Halliday & Webster (2009), while Halliday & Matthiessen (2013) offers a comprehensive description grammar as it is handled by SFL.

(i.e. contextual subsystems) of these metafunctions help to describe the kinds of meanings being expressed in a given utterance or text. Each metafunction works in parallel with the other two (Webster, 2009) and "contribute jointly to the description of linguistic units" (Hasan, 2014, p. 26). When considered together as a whole, these contextual elements form what is called register within SFL.

In the following subsections I will elaborate on what is meant by register in SFL, before going on to describe the individual metafunctions and contextual variables which register is made up of.

Register

Broadly speaking, register is a variation of any given language used for specific purposes and/or in specific situations. Perhaps the easiest way to 'visualise' the way register varies language is to imagine discussing a topic, say a recent assignment at your job, with your best friend, and then imagine how you would discuss that same topic with the CEO of your company. More likely than not, there would be some significant differences in how you go about using language in those two situations. In this sense, register can be seen as a kind of scale, ranging from the very informal (as when talking to a close friend) to the very formal (as when talking to an individual to whom a great deal of respect must be shown). It is important to emphasise that register variation is not limited to spoken language, but is a feature of all instantiations of language, the particularities of which will alter the expected register(s) used and how they are perceived. For example, e-mail communication tends towards an informal register (Posteguillo-Gómez, 2002), whereas university textbooks are largely more formal in terms of register.

In SFL, register is functionally motivated variation (subsystem) of a lan-

guage, and an amalgam of the contextual variables of FIELD, TENOR, and MODE. There is also the notion of register range or 'repertoire' which refers to the total range of registers available to a given individual or community (Webster, 2009, p. 246).

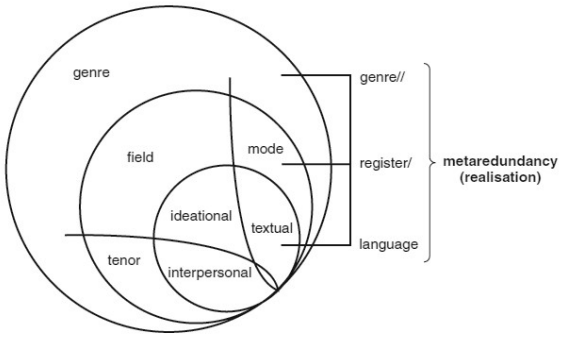


Figure 3.2: Register as it relates to metafunctions and contextual variables. Adapted from J. Martin (2009).

The Ideational Metafunction and the Contextual Variable of Field

The ideational metafunction is a composite of the 'experiential' and 'logical' metafunctions; the experiential relates to resources which enable individuals to exchange, i.e. share and discuss, their experiences with one another. These resources are involved in the encoding of the experience of the "internal and external world as processes of various kinds occurring in time and entailing entities as participants" (Hasan, 2014, p. 8). The experiential also plays a role in the elaboration of the circumstances surrounding how participant and process are configured, which is to say, the circumstances of the exchange which the participants find themselves in. A similar role is filled by the logical metafunction. This is a more focused metafunction in the sense that it does not relate to the circumstance as a whole, but is mainly concerned with

the relation between the individual configurations of participant and process. Furthermore, this function "elaborates on the nature of entities, processes, and circumstances by relating them in different ways to a set or properties" (Hasan, 2014, p. 8).

Field is related to the ideational metafunction in that it refers to, on the one hand, "the nature of the social and semiotic activity" (Halliday & Matthiessen, 2013, p. 33) taking place, and the domain of experience (i.e. subject matter) on the other, and in that way sets a limit on the focus of an interaction. Field itself consists of eight social-semiotic processes, which have been divided by Halliday & Webster (2009) into first- and second-order processes. The social process of doing, that is engaging in an action facilitated by language, is the only first-order process. The second-order processes consist of: expounding on an experience; reporting on an experience, event or region of space[12]; recreating experiences (including imaginary experiences, and other socio-semiotic processes); sharing values and/or experiences; recommending courses of action; enabling courses of action; and exploring values and hypotheses, including the comparisons of alternatives there of and arguing for or against them (Matthiessen, 2009, p. 31).

According to Butt (2004) as referenced in Wegener (2011), there are four major subsystems which constitute FIELD, namely: SPHERE OF ACTION, MATERIAL ACTION, ACTION WITH SYMBOLS, and GOAL ORIENTATION. The primary concern of the SPHERE OF ACTION subsystem is the need for the subject matter of a communicative instance to be defined. MATERIAL ACTION relates to the role played by activity in the Context of the given instance. That is to say, the role of non-linguistic activities in the overall communicative Context. The ACTION WITH SYMBOLS subsystem refers to the necessity, or lack thereof,

[12]Space here is used in the most general sense of a region, room, place or area, and not specifically in the sense of 'outer space'.

for language use in the completion of a given task within the a given Context. The fourth subsystem, GOAL ORIENTATION, seeks to map the motivations for action "as they are outwardly manifested" (Wegener, 2011, p. 150).

To provide a practical example of FIELD, let us briefly consider this subsection itself as a text. In order to keep the example brief I will limit my description of features to one level of delicacy. The SPHERE OF ACTION, i.e. what it is about, is *a description of ideational metafunction and the contextual variable of* FIELD. The MATERIAL ACTION is in this instance OBLIGATORY, because I, as the writer, must interact with a computer and keyboard to communicate with you, the reader. From your perspective as the reader, the relevant feature of MATERIAL ACTION would be OBLIQUE (MARGINAL) as the extent of required action in this instance is to accesses this text either in print or electronically. Given that this is a written text ACTION WITH SYMBOLS is NECESSARY, this is because I must use alphabetic characters in order to communicate. Because I am able to determine my own intentions for writing this text, GOAL ORIENTATION can be described, and in this case is OVERT + IMMEDIATE + CONSTANT; I am writing this example in order to explain FIELD to an individual who may be encountering this concept for the first time.

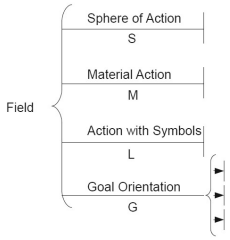


Figure 3.3: The contextual variable of FIELD and its major subsystems. Adapted from Wegener (2011).

The Interpersonal Metafunction and the Contextual Variable of Tenor

The interpersonal metafunction refers to the enacting of social and personal relationships through language (Halliday & Matthiessen, 2013, p. 30). This metafunction serves as a means for construing and interpreting social reality via the aforementioned social relationships. It also relates to the assessment and evaluation of possibilities and phenomena with regards to the roles of participants, one's attitude towards one's self and others, and "commitment to the interactive process" (Hasan, 2014, p.8).

As a means of identifying the Context of a given instance of communication, TENOR refers specifically to the participants. The roles played by the participants can be described in terms of their institutional roles, status roles (i.e. the power dynamics between the participants), contact roles (i.e. the familiarity or closeness of the participants), and sociometric roles (i.e. the affect of those involved, whether the neutral or charged, positive or negative). The values imparted by the participants also fall under the scope of TENOR (Halliday & Matthiessen, 2013, p. 33).

Tenor itself is made of up four subsystems. These are SOCIAL HIERARCHY, AGENTIVE ROLE, SOCIAL DISTANCE, and NETWORK MORPHOLOGY. SOCIAL HIERARCHY relates to the status and power relations between individuals engaged in a discourse, be they equal or unequal, mutable or immutable, whereas AGENTIVE ROLE directs focus towards the various roles of the individuals, as well as how those roles are established and maintained. The system of SOCIAL DISTANCE relates to the extent to which the individuals involved know each other. NETWORK MORPHOLOGY seeks to provide a framework for describing and relevant social network features of the discourse community as a whole.

As with the previous contextual variable, I will provide a brief example

of TENOR framed within this subsection as an independent text. Since I do not know exactly who I it is that may be reading this text, for the sake of example, I will interpret the various subsystems as though the reader is a professor in my department. With that in mind SOCIAL HIERARCHY is HIERARCHIC + DECLARED (EXPLICIT) + IMMUTABLE, this is because there is an explicit hierarchy to the relationship between professor and student, and in this exact instance these roles cannot change. The features of SOCIAL DISTANCE are: UNIPLEX, as relationship is limited to interactions within the department; REGULAR, because we interact at regular intervals, i.e. according to class schedules; there is also common CODAL SHARING and with regards to the being CODALLY DISTINCT strong framing and and classification of the relationship is present. AGENTIVE ROLE is ACQUIRED in this instance, as no one is inherently a student or professor of a given department. It is also CIVIC as opposed to FAMILIAL, and the roles involved are RECIPROCATING in nature, as a professor arguable requires a student and vice-versa. NETWORK MORPHOLOGY is both SCALAR in nature when factoring the size and diversity of the department, while NON-SCALE features are our POSITIONALLY DEFINED ROLES as student and professor, which are DYADIC as well.

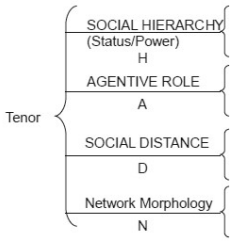


Figure 3.4: The contextual variable of TENOR and its major subsystems. Adapted from Wegener (2011).

The Textual Metafunction and the Contextual Variable of Mode

The textual metafunction contributes to the semiotic reality, that is the meaning-making reality, of a text or utterance and its creation. To put it another way, the textual metafunction relates to all of the resources available for creating and organising the communication of meanings, as well as the resources for cohesion and continuity as a communicative instance progresses (Halliday & Matthiessen, 2013).

MODE is a diverse variable which falls under the purview of the textual metafunction, and refers to the role "played by language and other semiotic systems in the situation" (Halliday & Matthiessen, 2013, p. 33), namely:

1. the division of labour between semiotic activities and social ones (ranging from semiotic activities as constitutive of the situation to semiotic activities as facilitating).
2. the division of labour between linguistic activities and other semiotic activities.
3. rhetorical mode: the orientation of the text towards field (e.g. informative, didactic, explanatory, explicatory) or tenor (e.g. persuasive, exhortatory, hortatory, polemic).
4. turn: dialogic or monologic
5. medium: written or spoken.
6. channel: phonic or graphic.

(Halliday & Matthiessen, 2013, p. 33)

Like the variables of FIELD and TENOR, MODE consists of several subsystems: ROLE OF LANGUAGE, CHANNEL, and MEDIUM. ROLE OF LANGUAGE focuses on the dominance and importance of language to the activity being undertaken, be it a supportive or constitutive role. CHANNEL seeks to describe the characteristics or a communicative signal, for example whether visual contact is involved, if the communication is graphic or phonic, and whether it is a monologue or dialogue (Hasan, 1999). The subsystem of MEDIUM is concerned with the medium of transmission, for example if the communication is written or spoken.

Once again, I will provide an example of which features of MODE of this subsection would be present, so as to provide a clear demonstration of it as contextual variable. Looking at the ROLE OF LANGUAGE it is CENTRAL to the communicative instance - in order to effectively make meaning in this situation I must use language. The CHANNEL used here is GRAPHIC as it is read and not heard, while the MEDIUM is WRITTEN not SPOKEN.

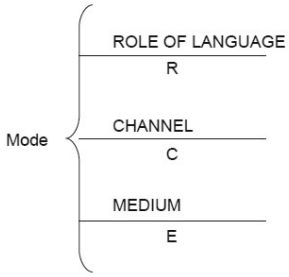


Figure 3.5: The contextual variable of MODE and its major subsystems. Adapted from Wegener (2011).

Having outlined the theory and foundational concepts of this work, the next chapter will be a detailed discussion of the data and methodologies used in this thesis, starting with a description of the sources of my data, before an explanation of the various methodologies which I made use of in my analyses.

4 The Data & Methodology

The data analysed here were initially collected by Jason Baumgartner of Pushshift.io as part of his Reddit Public Data Archive (hereafter: RPDA).[1] This is a publicly available archive comprised of all public Reddit comment and submission data posted by users to the site over a span of approximately nine years. Submission and comment data were collected separately and as such cover slightly different timespans; submission data are available from the period between January, 2006 to February, 2016, whereas comment data are available for the period spanning between October, 2007, and February, 2016.[2]

The RPDA consists of approximately 1.7 Billion[3] posts and represents digital data in excess of one terabyte when decompressed (Stuck_In_The_Matrix, 2015). Both comment and submission data are divided into monthly archives. The monthly submission data archives are in some cases incomplete, and certain periods were not recorded. This affects earlier years (i.e. 2006 & 2007) primarily, and does does not appear to represent any gaps in relevant data, as KW was first established in 2012. There are also gaps in the comment data as approximately 350,000 comments could not be collected due to issues with the Reddit API (Baumgartner, 2015). There are no data available as to what timespan may have been affected by this issue or if this affects archives across all months of the RPDA, or if this issue was limited to specific subreddits.

The individual archive files which make up the RPDA range in size from 13 megabytes to 6.3 gigabytes, based on the amount of redditor activity during

[1] The RPDA is not to be confused with the Archive hosted at www.redditarchive.com, which is affiliated with neither Jason Baumgartner and Pushshift.io, nor Reddit and its parent company Conde Nast Digital. This site simply hosts links to the most popular content and threads posted on Reddit, sorted by the date of posting and their cumulative Karma score

[2] As of writing, the archive appears to be continually updated, although no statements or information are available as to the intended duration of data collection.

[3] Estimate as of 17.07.2015.(Stuck_In_The_Matrix, 2015)

the collection period, as well as the nature of the data collected. The comment archives generally contain more entries per month, due to there often being multiple comments per submission. This results in larger average file sizes when compared to submission archives. Compressed, the monthly comment archives for the year 2015 average approximately 5.67 gigabytes in size compared to 1.75 gigabytes for the compressed monthly submission archives for the same period.

4.1 Dataset Source

Both the comment and submission data are hosted online at files.pushshift.io. As of the date of this writing, access to the data is available free of charge. The archive files making up the RPDA were compiled by Jason Baumgartner by accessing Reddit's API[4] (Stuck_In_The_Matrix, 2015). Access to the reddit API is provided to the general public free of charge, though they do have set rules for how the API is to be used and implemented by third-party programs. [5] For those who have the appropriate skill set, one can set about collecting the data contained within the RPDA manually, and build an archive which conforms to one's own set of specifications. Given the breadth of the RPDA, and the public access to these data available to other researchers should they wish to work with it,[6] I have chosen to make use of the RPDA for the purposes of this thesis.

[4] Application Programming Interface.

[5] As of writing, API access rules are hosted at: https://github.com/reddit/reddit/wiki/API

[6] Using publicly available datasets such as the RPDA allows for others to verify my claims should they wish to do so.

4.2 Data Collection & Preprocessing

As the RPDA was compiled by a third-party to include all available data from all public subreddits, a large amount of preprocessing was required in order to isolate the data relevant to this thesis and compile it into corpora (see: section 4.3). In the following section I describe the procedures used to collect and create these corpora. The same procedures were used for all subreddits within the corpora data, with the only differences relating to names of the various subreddits which were specified as necessary.

To facilitate the collection and preprocessing of the data relevant to this thesis, I made use of a private cloud-based Ubuntu server, as well as two open-source applications which enabled the direct input of commands and simple manipulation of files remotely from my personal computers. The Ubuntu operating system was suggested to me by Jason Baumgartner due to its built-in terminal interface, which in his opinion provided the least complicated method of manipulating the archives and extracting the relevant data (Baumgartner, 2016). The open-source applications used were PuTTY and FileZilla. "PuTTY is a free implementation of SSH[7] and Telnet[8] for Windows and Unix platforms, along with an xterm terminal emulator. It is written and maintained primarily by Simon Tatham"(Tatham, 2016). PuTTY was used as a means to provide input to, and execute commands in, the Ubuntu server. FileZilla is an open-source program which facilitates cross-platform file transfer. It was primarily used to transfer files between the server and my personal computers, and to monitor the progress of commands input through PuTTY. Strictly speaking, neither program was required; there are a number of alter-

[7]SSH stands for Secure Shell, and is a web-protocol which enables remote login services and data transfer over otherwise insecure networks.

[8]Telnet is a protocol which enables the bi-directional transfer of text-based communication across a network

natives which a researcher may use, should one wish.

The initial collection of the RPDA archive files was facilitated using the 'wget' command which enables the downloading of data via a number of different web protocols.[9] The command requests the transfer of data from a specified URL to the system acting as the issuant. However, wget does not modify data or otherwise interfere with its content or container format.[10]

After transferring the data, several steps were required in order to facilitate the analyses at the centre of this thesis. This is because of not only the format of the archival files themselves, but also the breadth and formatting of the data contained within. The RPDA consists of dozens of individual compressed data archives of the .bz2, or 'bzip2', compression format. This format is akin to other data-compression formats such as the more ubiquitous .zip format. However, unlike .zip, .bz2 is not natively supported within the Ubuntu environment. In order to manipulate these files, a third-party program, bzip2, was needed. This format and the extraction application are both open-source and can be used across a wide variety of operating systems.

The data within the RPDA archives is encoded in a format known as JSON. The acronym JSON stands for JavaScript Object Notation and "is an open-standard format that uses human-readable text to transmit data objects consisting of attribute-value pairs. It is the most common data format used for asynchronous browser/server communication"(Wikipedia, 2016a). Despite JSON data being classified as human-readable, which means that the contained data is written in ASCII or Unicode characters, the complexity of JSON formatted files can be such that it obscures the contained data and hinders immediate analysis. There may also be a great deal of metadata contained

[9]FTP, HTTP, and HTTPS are all currently supported.

[10]An example would be a command which downloads and extracts archival files formats such as .zip or .rar files.

within any given JSON-formatted file. The RPDA files contain not only this metadata, but also the complete collection of all user posts on Reddit over the period of the given month. This amounts to millions of lines of raw data which need to be processed prior to analysis. For this reason I executed several commands in order to parse through the files, looking for and extracting the relevant data, namely all posts on the r/kreiswichs, r/SCHLAND, r/de, and r/self_de subreddits. The command used for this stage of preprocessing was the 'grep' command. The command itself is an acronym of its function, and stands for 'global (search for a) regular expression and print (out matched lines)'. The script shown in figure 4.1 indicates the desired expression - in this case a string of characters which include the attribute "subreddit" and the name of the desired subreddit - and matches its value against the values which I have indicated. When the grep command finds a string matching the given value the entire contents of the line are output to a file of the indicated name and type.

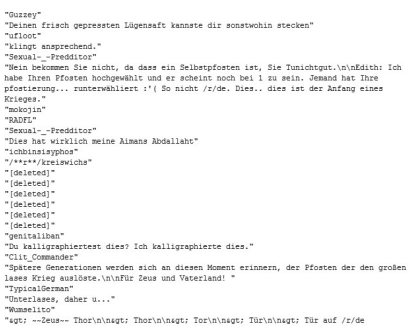
```
wget files.pushshift.io/comments/RC_2015-01.bz2 && wget
files.pushshift.io/comments/RC_2015-02.bz2 && wget
files.pushshift.io/comments/RC_2015-03.bz2 && bzip2 -cd RC_2015-01.bz2 | grep
'"subreddit":"de"' >> de-January-2015.txt && bzip2 -cd RC_2015-02.bz2 | grep
'"subreddit":"de"' >> de-February-2015.txt && bzip2 -cd RC_2015-03.bz2 | grep
'"subreddit":"de"' >> de-March-2015.txt && bzip2 -cd RC_2015-01.bz2 | grep
'"subreddit":"SCHLAND"' >> SCHLAND-January-2015.txt && bzip2 -cd RC_2015-02.bz2 | grep
'"subreddit":"SCHLAND"' >> SCHLAND-February-2015.txt && bzip2 -cd RC_2015-03.bz2 | grep
'"subreddit":"SCHLAND"' >> SCHLAND-March-2015.txt && bzip2 -cd RC_2015-01.bz2 | grep
'"subreddit":"self_de"' >> self_de-January-2015.txt && bzip2 -cd RC_2015-02.bz2 | grep
'"subreddit":"self_de"' >> self_de-February-2015.txt && bzip2 -cd RC_2015-03.bz2 | grep
'"subreddit":"self_de"' >> self_de-March-2015.txt && bzip2 -cd RC_2015-01.bz2 | grep
'"subreddit":"kreiswichs"' >> kreiswichs-January-2015.txt && bzip2 -cd RC_2015-02.bz2 |
grep '"subreddit":"kreiswichs"' >> kreiswichs-February-2015.txt && bzip2 -cd
RC_2015-03.bz2 | grep '"subreddit":"kreiswichs"' >> kreiswichs-March-2015.txt && rm
RC_2015-01.bz2 RC_2015-02.bz2 rm RC_2015-01.bz2 RC_2015-02.bz2 RC_2015-03.bz2
```

Figure 4.1: The script used to download and grep data from the RPDA into text files for inclusion in my corpora.

After the complete sets of data from the relevant subreddits were isolated into their own files, another series of commands were executed in order to strip the unneeded non-linguistic metadata from said files. The commands involved

{"author_flair_text":null,"gilded":0,"id":"co781d3","score_hidden":false,"parent_id":"t3_
2ub16p","ups":2,"distinguished":null,"downs":0,"created_utc":"1422749936","author_flair_c
ss_class":null,"name":"t1_co781d3","body":"Deinen frisch gepressten L\u00f6gensaft
kannste dir sonstwohin
stecken","subreddit_id":"t5_2vlhq","link_id":"t3_2ub16p","controversiality":0,"author":"G
uzzey","retrieved_on":1424281506,"edited":false,"subreddit":"kreiswichs","archived":false
,"score":2}
{"subreddit":"kreiswichs","archived":false,"retrieved_on":1424281158,"edited":false,"auth
or":"ufloot","score":2,"link_id":"t3_2ub29a","subreddit_id":"t5_2vlhq","controversiality"
:0,"downs":0,"ups":2,"distinguished":null,"author_flair_css_class":null,"name":"t1_co78s6
s","body":"klingt
ansprechend.","created_utc":"1422751384","gilded":0,"author_flair_text":null,"parent_id":
"t1_co76cpa","id":"co78s6s","score_hidden":false}
{"author_flair_text":"Mohzhammed","gilded":0,"score_hidden":false,"id":"co7906z","parent_
id":"t3_2ucxam","distinguished":null,"ups":22,"downs":0,"created_utc":"1422751819","body"
:"Wein bekommen Sie nicht, da dass ein Selbstpfosten ist, Sie Tunichtgut.\n\nEdith: Ich
habe Ihren Pfosten hochgew\u00e4hlt und er scheint noch bei 1 zu sein. Jemand hat Ihre
pfostierung... runterw\u00e4hliert :'(So nicht /r/de. Dies.. dies ist der Anfang eines
Krieges.","name":"t1_co7906z","author_flair_css_class":"","subreddit_id":"t5_2vlhq","link
_id":"t3_2ucxam","controversiality":0,"retrieved_on":1424281054,"edited":false,"author":"
Sexual-_-Predditor","subreddit":"kreiswichs","archived":false,"score":22}
{"author_flair_text":null,"gilded":0,"score_hidden":false,"id":"co79ris","parent_id":"t1_
cnx8qc3","distinguished":null,"ups":2,"downs":0,"created_utc":"1422753363","name":"t1_co7
9ris","body":"RADFL","author_flair_css_class":null,"subreddit_id":"t5_2vlhq","link_id":"t
3_2tanZs","controversiality":0,"author":"mokojin","retrieved_on":1424280700,"edited":fals
e,"subreddit":"kreiswichs","archived":false,"score":2}
{"score_hidden":false,"distinguished":null,"subreddit":"kreiswichs","edited":false,"gilde
d":0,"retrieved_on":1424280371,"controversiality":0,"link_id":"t3_2ubbxu","subreddit_id":
"t5_2vlhq","ups":4,"id":"co7ac7w","author_flair_css_class":"","downs":0,"parent_id":"t3_2
ubbxu","author":"Sexual- -Predditor","score":4,"name":"t1_co7ac7w","body":"Dies hat

(a) RPDA data prior to being parsed using jq



"Guzzey"
"Deinen frisch gepressten Lügensaft kannste dir sonstwohin stecken"
"ufloot"
"klingt ansprechend."
"Sexual-_-Predditor"
"Wein bekommen Sie nicht, da dass ein Selbstpfosten ist, Sie Tunichtgut.\n\nEdith: Ich
habe Ihren Pfosten hochgewählt und er scheint noch bei 1 zu sein. Jemand hat Ihre
pfostierung... runterwähliert :'(So nicht /r/de. Dies.. dies ist der Anfang eines
Krieges."
"mokojin"
"RADFL"
"Sexual-_-Predditor"
"Dies hat wirklich meine Aimans Abdallaht"
"ichbinsisyphos"
"/**r**/kreiswichs"
"[deleted]"
"[deleted]"
"[deleted]"
"[deleted]"
"[deleted]"
"[deleted]"
"genitaliban"
"Du kalligraphiertest dies? Ich kalligraphierte dies."
"Clit_Commander"
"Spätere Generationen werden sich an diesen Moment erinnern, der Pfosten der den großen
lasea Krieg auslöste.\n\nFür Zeus und Vaterland! "
"TypicalGerman"
"Unterlasea, daher u..."
"Wumselito"
"> --Zeus-- Thor\n\n> Thor\n\n> Tor\n\n> Tür\n\n> Tür auf /r/de

(b) RPDA data after using jq

Figure 4.2: Two versions of the same JSON data taken from the RPDA from the month of February, 2015

were the "cat" command, which stands for concatenate, and in this instance was used to call up a file and feed its contents into a following command, namely, the "jq" command. The jq program and its associated command is a freeware program designed to allow a user to filter JSON files for a given set of objects, whose values are then output to a new file if desired. In my analysis the author and post content is of most relevance, and so the jq commands were written in such a way as to search for and output all data matching the '.author' and '.body'[11] headings within my stripped down files. The results were then output into plain-text files which were given names based on the subreddit from which the data come and the month in which they were posted. This same method was used to obtain files with only the body of posts (submission and comment data) but without the corresponding author data. These data were used for quantitative analysis, where the author of a given post is of little relevance. All of the above mentioned processes create separate output files leaving the original data intact should it be needed.

The above process was also used to create plain-text files of submission[12]

[11] '.body' in this case refers to the 'body' or content of a given post.

[12] In this context 'submission' refers to what most would call a thread or post.

data. There were some minor differences, however, which resulted from the separate collection and hosting of the submission data archives within the RPDA as mentioned in section 4.1. Aside from accessing a different directory during the downloading process, the only significant change to the script used was the replacement of the object '.body' (see above) with '.title'. The .title object refers to the title of a submission.

The plain-text within the files was then run through a plain-text conversion tool in order to format the text. This was done to remove vestigial elements of the JSON formatting which may not be properly recognised within various analysis software. The data removed were the character combination '\n', which is a common method to denote a 'new line' in the JSON formatting[13]. While this tool reformatted new lines and most ASCII characters properly, retaining them for the output file, it did not convert the HTML entities such as '>' and '<' which denote the characters '>' and '<', respectively. The characters are critical to understanding larger ASCII based images as well as the formatting and associated meanings of certain texts. To ensure that these characters were present in the final files, the 'search and replace' function of Notepad++ was used.

The data remaining within the plain-text files after preprocessing represents all comment and author data from the indicated subreddits over the course of the year 2015. The data have been grouped into several sets of files, first sorted by subreddit and month, then by comment or submission data. Of these sets, there are two versions (A & B). For comment data, version A contains author and body data, while version B contains only the comment body. For submission data the files are similarly divided, with version A containing

[13]This character set tells a system to display the following text on a new line, and is the equivalent to what a user would see upon hitting the return key while using word processing software

author and title data, and version B containing the submission title. A master file of each set and version has been created, with a compilation of all data. There is also a merged master file with both comment and submission data sets. Aside from the aforementioned objects, neither the individual monthly files nor master files contain any further metadata.

The order of text in the individual monthly files has been unaffected by the various parsing and output commands, and retains the original order of the non-manipulated archival files. This means that all lines of textual data from both the comment and submission archives are in the chronological order in which they we collected during the creation of the RPDA. This order is based on the UTC[14] timestamp metadata as assigned at the time of a comment's posting. The master files retain this order with the exception of those with both comment and submission data. These files, which were created for the purposes of quantitative analysis, contain comment data in chronological order followed by submission data in the same order. This is due to technical limitations which inhibited the collation of the data. The issue specifically relates a lack of any specific metadata which can reliably link submissions and comments. Using UTC timestamps would provide mixed results at best, as submissions and comments from different users can be posted at the same time, and the comments within a submission may come long after new submissions have been posted causing a UTC-based sorting to align a comment with the wrong submission. Ultimately, this does not impact this work because the analyses therein do not seek to create a chronological account of KW's development.

[14] 'Universal Time, Coordinated'.

4.3 Corpora

The preprocessing described in 4.2 was used to create the two corpora which form the backbone of this thesis' data. The corpora are divided into two sets of data. The KW-corpus was created from data from the kreiswichs subreddit. The reference corpus, the 'German Reference Subreddits' (GerRefSr) corpus, includes data from the r/de, r/SCHLAND, and r/self_de subreddits; a reference corpus serves as a point of comparison for data found in a primary, or source, corpus. Reference corpora allow researchers to make determinations about which data may be representative of unique phenomena based on their relative frequencies of occurrence. If a lexical item is relatively more frequent in a source corpus when compared to an appropriate reference corpus, it may represent an item requiring closer investigation.

Both the Kw-Corpus and GerRefSr contain alternate sets of files of both comment and submission data. These datasets are in separate plain-text files and include two versions[15] of the comment and submission data. Version A includes both author and body/title data, while Version B contains only the body/title data. In the following subsections I will provide a quick overview of the sizes of these corpora.

The KW-Corpus

The central corpus for this thesis is the KW-Corpus which was compiled through the procedures laid out in section 4.2. Although the post history on KW goes back to the creation of the subreddit on November 20, 2012 (Redditmetrics, 2016b), the most complete archives for both KW and the reference subreddits are from the year 2015. For this reason the KW-Corpus and

[15]This is due to the simplistic encoding of the source files. For the quantitative analysis, the author of a post was irrelevant, and thus separate files without said data was created.

the reference corpus contain only data for that year.

The KW-corpus version A contains 74,545 lines of textual data, of which 27,646 lines are the name of the author[16] of a given post, with the remaining 46,899 lines representing comment and title data. This is an average of 6,212[17] lines of text data per month. The median length is 5,838 lines of text. February, 2015, contains the least data at 4,466 lines, whereas October, 2015, represents the largest dataset at 8,648 lines[18].

Version B is 46,899 lines of data in size. Again, February is the month with the least amount of data recorded, with 2,715 lines, while August has the most, with 5,595 lines. The median length is 3,397 lines, and the monthly average is 3,908[19] lines of text data over the 12 month period.

The corpus consists of 23,775 comments spread across 3817 submissions. This is an average of approximately 6 comments[20] per submission.

The GerRefSr-Corpus

The GerRefSr corpus was created to serve as a point of comparison for this thesis. The lexical items found in the GerRefSr represent the standard lexical items of the German-language community of Reddit. As mentioned in section 4.3, This corpus consists of data from the following subreddits: r/de, r/SCHLAND, and r/self_de.[21]

[16] All usernames require only one line in order to be displayed, but depending on how a post is formatted, comments can span several lines. This is the case when a user makes use of the return key to create a new line, whereas simply ending a sentence with a period does not trigger the creation of a new line.

[17] Rounded down from 6,212.083.

[18] These figures were obtained using the line number listing feature Notepad++ program, which in essence displays the number of times that a new line is called up in a given file via either a newline delimiter or use of the return key. Other software may calculate lines differently.

[19] Rounded down from 3,908.25.

[20] This figure was rounded down from 6.228.

[21] For a description of each subreddit, see chapter 2 section 2.2.

GerRefSr contains substantially more data than the KW-Corpus. This is due to the relatively large and active userbase on the r/de subreddit, which can be seen in table 4.1.

Subreddit	Version A		Version B	
r/de	1,212,724	(99.52%)	852,219	(99.59%)
r/SCHLAND	5,557	(0.46%)	3,301	(0.39%)
r/self_de	278	(0.02%)	173	(0.02%)
GerRefSr	1,218,559		855,693	

Table 4.1: The constituent subreddits of the GerRefSr corpora, the total lines of text which they contribute and the percentage thereof relative to the corpus as a whole.

Version A of the GerRefSr corpus consists of 1,218,559 lines of text data, while Version B contains 855,693 lines. Versions A and B represent 101,547[22] and 71,308[23] lines of data on average per month. In terms of monthly data, February contains the least amount for Version A with 52,751. For Version B, April represents the smallest dataset with 31,410 lines. The dataset for the month of November was the largest for both versions of GerRefSr with 183,805 and 132,554 in Versions A and B, respectively. The median amount of data was 107,541 lines in Version A and 53,149 in Version B.

4.4 Methodology

Both quantitative and qualitative methods were used to analyse the data contained within the KW-corpus (see: Section 4.3). First, a quanitative analysis is conducted to establish if, and if so, to what extent, the KW-corpus data differs from that present in the GerRefSr. This process involves an accounting of the lexical items present in KW-Corpus-B.[24] I should note that the

[22]Rounded up from 101,546.583.

[23]Rounded up from 71,307.75.

[24]Version-B of the corpora were used as I did not wish to include usernames in these data.

term *lexical item*, rather than *word* is used through the following sections and chapters as it more clearly reflects the data being analysed; online discourse and meaning making often includes not only words as they are commonly understood, but also often numbers, individual ACSII characters, and images and items created using combinations thereof. After conducting a quantitative analysis, the lexical items which occurred at a relatively high frequency within the KW-Corpus compared to the GerRefSr-Corpus were then chosen for qualitative analysis, which was undertaken with the goal of uncovering the potentially unique or novel features of these items. After investigating the lexically significant aspects of KW, my analysis expands in scope to explore the discourse of KW at the level of individual user submissions. This was done to identify the contextual elements of discourse (see subsections: 3.2, 3.2, and 3.2) which play a particularly strong role within KW as a discourse community, and which elements are either encouraged, discouraged, or even shunned.

In the following sections, I will first provide a quick definition of various lexical items which are looked at in this thesis, followed by detailed account of the methodologies used in the course of my analyses.

Lexical Items

Given the subject matter of this thesis and the nature of online discourse, the most appropriate units of meaning making for me to consider are, as have briefly mentioned above, not words, *per se*. This is because other non-word items carry and impact meanings in similar ways to purely alphabetic words in texts or intonation and gesticulations in spoken communication. Consider for example the following brief exchange featured in figure 4.3.

There are several instances whereby non-word items contribute significant

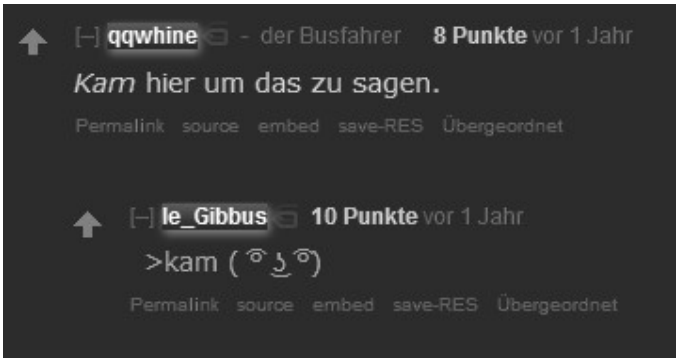


Figure 4.3: A brief exchange between two users with significant non-alphabetic lexical items.

meanings to the communication. First there is the method for highlighting specific words within the text, namely the greater than symbol (in this case used in an arrow-like capacity to direct attention). The italicisation is also incorporated by user qqwhine[25] in order to draw attention to the word kam and its potential double meaning. User "le_Gibbus" acknowledges this double meaning by again highlighting the same word, this time using >. This is then followed by the so-called 'lenny face' (NovaXP & Asdfghjkl., 2012) to further reinforce the innuendo of the statement. Without the use of non-word items, the meanings intended by both parties would be more ambiguous and open to interpretation. To that end they are included in the analyses of this thesis.

The following list provides context to the four types of linguistic items which I have identified for the purposes of my analyses described in the following subsections:

1. **Alphabetic**: These are lexical items which are primarily, but not necessarily alphabetic in nature e.g. '90er-Kinder', /l/de, [entfernt], Aurum.

[25]This user's name also makes meanings with the use of non-word items. The 'qq' in his or her name is in fact an emoticon of a crying face, whereby the tails of the 'q'represent streams of tears coming from the eye, which is represented by the loop of the 'q'. This becomes more visually clear when one includes a third character to represent a mouth. e.g. q_q . The inclusion of 'whine' in the username further underscores the meaning.

2. **Numeric**: This includes those items which are soley or primarily numeric characters e.g. 11/9, 88, 16:20.

3. **ASCII Images**: These include lexical items which contain solely non-alphabetic characters (e.g. <—-) or which contain alphabetic characters solely for the purposes of completing an image (e.g. _o)>).

4. **Image Macros**: These are lexical items representing strings of text which are used to display specific, non-ASCII based images in KW posts. ex: [](/kant), [](/rarpeter1).

Quantitative Methods

In order to conduct the qualitative analysis, I made use of corpus analysis software developed by linguist Laurence Anthony,[26] particularly his program AntConc version 3.5 (Anthony, 2016). AntConc is a piece of software which includes several tools for the analysis of corpora, including wordlist generation, displaying and plotting concordances, keywords in context (KWIC), and collocations. For the quantitative portion of this thesis, I used AntConc primarily to generate ranked wordlists, both with and without frequencies.

For my quantitative analysis, the most relevant feature of AntConc was its keywords list generator. This was used to generate an ordered list of so-called keywords[27] which appear in my KW-corpus. This is done by comparing the relative frequencies by which words appear in one corpus (the KW-Corpus), against the frequencies of those same words in a second, 'reference' corpus (the GerRefSr-Corpus). Words which appear with a higher frequency in the first,

[26] Anthony is a prolific writer of software which enable various kinds of corpus and textual analyses. As of writing, he offers a number these programs, free of charge on his website at www.laurenceanthony.net

[27] As described in the introduction to section 4.4 a keyword is not necessarily a "word" in the colloquial sense.

or 'source', corpus are correspondingly assigned a higher KEYNESS value, than those words with a frequency of occurrence more similar or lower than that of their appearance in the second corpus. KEYNESS values are not capped at any particular value and many be either positive, as in the scenario previously described, or negative. In the case of negative KEYNESS, the value indicate a keyword which is more common in the reference corpus than in the source corpus.

While KEYNESS is itself a valuable tool for determining lexical difference between two corpora, it must be supplemented with another value in order to have true descriptive power. The value in question is EFFECT, which can also be calculated using the keyword function of AntConc. An EFFECT value is a measure of "magnitude of a observed finding" (Rosenfeld & Penrod, 2011), which in this case a KEYNESS value, and ranges from 1 (complete significance) to 0 (complete insignificance). EFFECT quantifies the likelihood that the calculated KEYNESS value would remain the same if both corpora were of the same size. Taking EFFECT into account is important as it is "a measure of the practical significance of a result, preventing us claiming a statistical significant result that has little consequence" (Ridge & Kudenko, 2010). To put it another way, EFFECT is useful metric because it allows those conducting a study corpora of uneven size to be certain of a result, as is the case in with the corpora used in this thesis. Conversely, without taking EFFECT into account one it would be much more difficult to assert that the calculated KEYNESS of keywords in fact represents meaningful data for analysis.

AntConc 3.5 offers users a number of different methods for calculating KEYNESS, and EFFECT, as well settings for thresholds of statistical significance (p-value). The algorithm which I used for calculating KEYNESS was the log-likelihood; p-value ≤ 0.05. I chose to calculate EFFECT using the difference

coefficient algorithm developed by Leech & Fallon (1992), as it was specifically designed to account for differing frequencies of occurrence of words in corpora of uneven size. For the purposes of this process all data were treated as lower-case. This was to help mitigate the effect of variations of capitalisation, intentional[28] or otherwise, on the results.

Accounting for URLs

It should be noted that regardless of the settings selected for generating keyword lists, a word list for the primary corpus must also be generated. As online discourse often incorporates a wide range of characters beyond traditional letters to communicate meanings, it can prove difficult to generate coherent word frequency lists which include all relevant data. To that end, based on the discourse already casually observed on KW, I developed a list of characters which were to be calculated as valid for the purposes of my word list. As my character list did not include the common punctuation markers, I encountered an issue whereby web URLs would be heavily divided, resulting in a great deal of interference in my word lists, and therefore keyword list. In order to minimize this, I chose to remove as many URLs as possible[29] from version B of both the KW-Corpus and the GerRefSr, and create a URL free version to be used solely for this aspect of the quantitative portion of this theses; the URLs remain intact in all other data.

[28]It is not uncommon for capitalisation to be used and manipulated for creative effect in online discourse, particularly when users are SPing (see: 2.4)

[29]The URLs were removed using commands similar to how the corpus itself was generated. The script targeted and deleted all strings of text which began with the character combination 'http'. Due to the very wide range of possible URLs, some did escape this process and can be found in the data used for calculating keywords.

Qualitative Methods

As mentioned in the introduction to this section, the quantitative analysis was focused on establishing lexical differences between the KW and GerRefSr corpora. The differences discovered within the KW-Corpus were then analysed using qualitative means in order to discern if they correspond to the features of an anti-language as laid out by Halliday (see section 3.1). While the lexicon of a discourse community is quite an interesting and illuminating element of analysis, it does not, on its own, provide a comprehensive picture of a community's discourse. To that end, I made use of qualitative discourse analysis methods as well.

Before describing the methodology employed at the discourse level, I will elaborate on the qualitative analysis of lexical items. After identifying items of interest via quantitative means (see subsection: 4.4 above) as well as those found within the KW *Begrüßungsfaden* (see section 2.3), I analysed the list, looking for novel items, or items which appear somewhat out of place in online discourse. Upon identifying such an item, I attempted to determine its origin to the greatest extent possible.[30] This included comparing items against a glossary of terms provided by the KW community in said *Begrüßungsfaden*. In the case of unlisted items which were potential loan translations, I translated the terms into English[31] myself. All of the items, both in their untranslated and translated forms, were then investigated for any listings in databases of internet terminology, slang, and memes. The primary sources used for this process were the websites UrbanDictionary.com, KnowYourMeme.com,

[30]The etymology of internet slang is currently an underdeveloped area of academic pursuit, both as result and source of difficulty investigating the appearance of an item.

[31]I admit that there is a presupposition on my behalf that many novel lexical items are of English origin. This is due to the overwhelming influence of the English language on Reddit, as can be seen in the glossary of translations provided by the members of KW themselves, and the discourse of the internet more generally.

Sprachnudel.de, and Mundmische.de as they represent some of the most comprehensive repositories of these data with sources and timestamps.

Moving on to discourse analysis, in order to restrain the scope of this thesis, I chose to analyse the register (see section 3.2) of user submissions as the centre piece of this thesis. An SFL-based conception of register is broad enough to provide a well-rounded understanding of the subject, while constrained enough so as to allow an in-depth discussion which does not exceed the scope of the present format, i.e. a Master's thesis. In order to accomplish this I made use of the framework established within SFL, using systems of features to map out the variations of TENOR, FIELD and MODE as they appear in each submission. This process is a straight-forward one, whereby a submission is analysed against the backdrop of each contextual variable. Let us take, for example, a hypothetical submission which makes reference to Aiman Abdallah. Such a reference is describable as a feature of the SPHERE OF ACTION subsystem of the contextual variable FIELD (see appendix for the full system networks). To map this, we work our way through the system, making note of the features which must be present in order for a reference to Aiman to be realised.[32] We then move on to the other (sub)systems and variables until we have a complete map of the submission's register. A total of 55 submissions were analysed using these methods. As will be discussed in greater detail in the chapter 5, I adopted, and in some cases adapted, the systems found in Wegener (2011).

I should at this point elaborate as to why and how submissions were in fact selected for analysis.

[32] In this case the chain of features would be: QUOTIDIAN → INSTITUTIONAL → PRACTICE → GENERAL → ROUTINE → ESTABLISHED → MEME → PUBLIC FIGURE → AIMAN ABDALLAH

Selecting Submissions for Analysis

Due to my interest in studying the KW-community framed within the concept of anti-languages and their related anti-society, it struck me as relevant to find the "edges" of their discourse. That is to say, the Contexts whereby a potential member is either embraced or excluded as far as could be determined by the individual's attempt at engaging in KW-discourse. Since in an online environment, the most definitive rejection of an individual attempting to make contact with a group is to simply erase their post (be it a comment or submission), I was presented with a challenge: I cannot analyse what I can no longer access. The next best thing would be to find the submissions of individuals who have only submitted content once to KW; single submissions being suggestive of a potentially failed attempt at integration and the Context surrounding it. My focus was placed on submissions rather than comments, because they generally represent a greater investment on the behalf of an individual in terms of engaging with an online community. For the year 2015, there were 94 submissions which met those criteria. The data used for this initial selection of individual submissions were obtained from the RPDA, however, the submissions (i.e. the submissions *in situ*) themselves were obtained directly from Reddit's own submission archives. This was done to guarantee that user evaluations of the submissions were as current as possible, since submissions are archived six months after being submitted. However, until such time they may be voted and commented on by users, and it is unclear exactly how long after a submission is submitted, that it is integrated into the RPDA. Also, in the RPDA, comments and submission are archived separately, which could lead to complications while attempting to correctly connect submissions to their associated chains of comments and replies. After retrieving the submissions, their relevant statistics (i.e. submitter, title, karma score, percentage

of upvotes, number of commenters, language used, etc.) were recorded and compared. During this process eight submissions were eliminated from potential analysis. Seven were stricken due to the deletion of their associated user account, which prevented me from determining if a user did in fact submit content to KW prior to, or after, the calender year 2015. A further submission was eliminated when it was discovered that the user had posted subsequent submissions in 2016. This process resulted in a total of 86 submissions suitable for potential analysis.

After compiling my data, I proceeded to narrow the focus of my dataset by sorting the submissions in order of their upvote percentage from highest to lowest. All submissions with 100% approval (i.e. no downvotes) were selected for closer analysis. The total number of submissions with this rating was 11. I then selected the bottom 11 submissions according to the above mentioned statistic. For this group of submissions, the approval percentages ranged from 25% to 68% at the time of their initial collection from Reddit.[33] This resulted in an initial total of 22 submissions for qualitative analysis.

In order to get an even clearer view of the kinds of submissions endorsed and disapproved of by members of KW, it would be pertinent to retrieve a comparable set of data which focus on submission karma score regardless of the user's posting history. To that end I retrieved the top 12 submissions of all time based on total Karma score, as well as 12 submissions with zero net karma.[34] Of the 12 zero-karma submissions, three were already retrieved as part of of the group of low approval submissions by one-time submitters. This resulted in a total of nine new unique submissions with a karma score of zero. I also retrieved a control group of 12 submissions selected at random.

[33]For an explanation as to how Reddit's karma system works, and why the submission karma scores may vary please refer to subsection 2.1.

[34]Submissions cannot be voted in to negative karma, though downvotes are still registered

The random selection was done using the following method: Using RES (see section 2.5) I loaded as many KW submissions as possible from the calender year 2015, sorted by top (i.e. karma) score. This resulted in 32 'pages' of submissions. Then using a random number generator,[35] set to select numbers from two to 32.[36] This set of numbers was then used to determine which pages I would select submissions from. A second set of 12 numbers between one and 21 was then generated to determine which submissions would be selected from the previously ascertained pages. Submissions were then chosen by counting down from the first submission listed on for each page until reaching the randomly determined number. These three collection processes resulted in a 33 additional submissions for analysis bringing the total to the previously stated 55.

Mapping Submission Features

Given that there are no pre-existing system networks yet developed for use in mapping Reddit-based discourse, it was necessary to create several for that purpose. In order to accomplish this, I modified two system networks found in Butt (2004) (as presented in Wegener (2011)). The modifications made are extensions in delicacy (see chapter 1 section: 1.5) of the ACTION WITH SYMBOLS and SPHERE OF ACTION subsystems present within the contextual variable FIELD (see chapter 3, section 3.2).

The various features and added layers of delicacy added to the above mentioned (sub)systems were determined based on personal observations of KW- and Reddit-discourses. For that reason, my system layouts should be consid-

[35]The generator used was the online random number generator service available at random.org. For more information as to how their generator works, please refer to their homepage.

[36]The first page was excluded as it contained the all the submissions already collected as part of the group of top karma submissions

ered tentative, and open to further development. It should also be noted that the arrays of features at the ends of these systems are those which are most emblematic of KW and of relevance to this thesis. However, by no means do they represent the full extent of possible choices. To view the diagrams of all system networks discussed in this thesis, please refer to the appendix.

Now that the methods used to conduct my analyses have been described, it is time to present their results.

5 An Analysis and Discussion of the Results

As discussed in chapter 4, two sets of methods, a quantitative and a qualitative were employed to make sense of the datasets described in chapter 4. The quantitative methods focused primarily on establishing whether or not my hypothesis that the lexicon of KW differs from that of the GerRefSr-Corpus. As such it is limited in focus to the lexical items collected into a single corpus as outlined in section 4.3 of this thesis. My qualitative methods are broader in scope as they were employed to analyse not only individual lexical items, but user discourse in the form of submissions as well. In the following sections and subsections, I will outline the results of these analyses, beginning with the quantitative, then moving on to the results of the qualitative analyses, first of the lexical items, which in turn is followed by and analysis of KW discourse.

5.1 A Quantitative Analysis of the KW Lexicon

Using the methods described in subsection 4.4, the KW-Corpus contains 374,055 tokens, of which 45,941 are unique lexical items. Of those 45,941 lexical items, I have uncovered a total of 1648 keywords with a positive KEYNESS value, when comparing the KW-Corpus-B against GerRefSr-B[1]. This equates to approximately 3.59% of all KW-Corpus-B lexical items. The total number of tokens recorded for these items is 133,367 which represents just over 35.65% of all tokens in KW-Corpus-B. What this means, in plain terms, is that despite amounting to less than 4% of all unique lexical items, these keywords make up more than one third of all items used in KW discourse.

When placing focus on the keyword measure of statistical significance, EFFECT, we encounter 1,549 lexical items with a significance above 50% (0.5000).

[1]These data exclude URLs as described in subsection 4.4.

These items constitute 3.37% of all unique items, and 96,594 (25.82%) of all tokens within the KW-Corpus-B. While this is a large reduction of the overall token count compared with the total of all keyword-tokens, the general result is quite similar. A small portion of statistically significant items, that is just over 3% of all items, represent a substantial portion of the data as a whole.

Returning to all items identified within by the keyword algorithm, the average KEYNESS of the items was 304.11[2] (median: 46.04; maximum: 213,078.29; minimum: 23.79), while the average EFFECT was 0.8800[3] (median: 0.9619; maximum: 1; minimum: 0.0956). When sorted according to their EFFECT value, the average KEYNESS of the top 25th percentile of keywords is 732.34 (median: 45.54; maximum: 213,078.29; minimum: 27.7), while their average EFFECT value is 0.9996 (median: 1; maximum: 1; minimum: 0.9939). For the bottom 25th percentile the average KEYNESS is 108.86 (median: 45.03; maximum: 2146.78; minimum: 23.82) with an average EFFECT of 0.6237 (median: 0.6840; maximum: 0.8410; minimum: 0.0956).

Looking at these data, it becomes apparent that there is a keyword whose impact on these quantitative data cannot be ignored. This is due to its incredibly high frequency of occurrence and KEYNESS. The lexical item in question is */r/wumselitospieleforum* (hereafter: WSF), which occurs 30,497 times in the KW-Corpora. This item alone accounts for just over 8.15% of all tokens in KW-Corpus-B, and just under 22.87% of all keyword tokens. As such it has an incredibly high KEYNESS value of 213,078.29 (EFFECT: 0.9999). This value is over 12 times greater than that of the item with the next highest KEYNESS, *fap* (KEYNESS: 17117.17; EFFECT: 0.9995). The term itself is the same of a subreddit associated with KW, and in particular the highly active

[2]All KEYNESS values have been rounded to the nearest hundredth.
[3]All EFFECT values have been rounded to the nearest fourth decimal place.

KW member 'Wumselito'. As of writing, his account has been deleted[4] he was both a moderator of KW and among the most recognisable posters across the German-language Reddit community. The subreddit itself is a reference to the Reddit 'Warlizard Gaming Forum' meme (kharanos & noob, 2013). The explanation for the incredible frequency of this item is the use of it in a copypasta (see section: 2.4) which consisted of 412 back-to-back instances of the term. This copypasta was then used several dozen times afterward, and represents all but nine instances of the item[5] Copypastas of this nature account for several high frequency items including: fap, telekom, 1945, [](/harald), selternerpeter, among others, though none are nearly as frequent as the aformentioned WSF.

Taking the WSF copypasta into account, there is a noticeable effect on the overall data relating to proportion of keywords to general lexical items in the KW-Corpus-B. The adjusted figures are: 343,567 total tokens, 29.94% (102,879) of which are keyword tokens (Avg. KEYNESS; 174.82; Avg. EFFECT: 0.8794). This is a relative decrease of approximately 17.02% from the original figure of 35.65% stated above, or in raw terms, a drop of 5.71%. This is certainly not an insignificant adjustment, particularly when looking at the impact it has on average KEYNESS which as been nearly halved from the previous figure 304.11. However I would argue that the adjustment is not necessary for the purposes of the present thesis. This is because these instances of WSF are representative of several legitimate internet phenomena in general, particularly SPing, copypasta, CJ-discourse, and of KW in particular. As such I have decided to include it in all further quantitative data.

For the purposes of the qualitative analysis (see subsection 5.2 below), I

[4]The relevant data still remain in the RPDA, as the archives in question were collected prior to the deletion of his account, however if you were to come across his posts now, the posters name is now [deleted], as are all posts linked to deleted accounts.

[5]the copypasta occurs 74 times in total, 73 of which were by a single user named 'IKraftI'.

selected the top 1000 keywords, sorted according to their EFFECT. This is a large number of items, however, it was the best way to ensure a reasonable breadth of keywords to investigate. The reason for this is how AntConc sorts lexical items. Regardless of which value is used to sort initially, any items with equal values are then displayed in order of the character's encoded value rather than alphabetically per a Latin-based alphabet, for example. Since I wanted to ensure that I did not only look more closely at items beginning with any specific set of non-letter characters, and because my data in not lemmatised resulting in different forms of the same word being treated as unique, 1000 was deemed to be reasonable. As all of the keywords have a positive KEYNESS value, I placed emphasis on their EFFECT value instead, which is to say their statistical significance. The relevant lexical items can be broken down as follows:

Item type	# of unique items	# of Tokens	Avg. Tokens per Item	Avg. KEYNESS	Avg. EFFECT
Alphabetic	830 (83.0%)	56074 (83.99%)	68	435.66	0.9806
Numeric	33 (3.3%)	1733 (2.60%)	53	323.56	0.9895
ASCII	82 (8.2%)	7421 (11.11%)	90	552.65	0.98655
Image Macros	54 (5.4%)	1534 (2.30%)	28	196.72	1
total	1000	66763	67	428.66	0.9824

Table 5.1: The top 1000 KW keywords by type and quantity.

As these data show, Alphabetic keywords dominate in terms of diversity and number of tokens. However, looking at the KEYNESS EFFECT values, it becomes clear that all four types of lexical item play a role in KW discourse and are significant. It also appears that there is an inverse relationship between KEYNESS and EFFECT, meaning that the more key a word is, the less one can be assured of the statistical significance of the result. This can be explained by the total number of the keywords. The more unique items there are, the greater the chance any one of them may also be found in the GerRefSr-corpora.

This in turn impacts the average EFFECT values, pushing them downwards. Meanwhile high frequency, and thus high KEYNESS items, .e.g. WSF, and 'fap', exert upward pressure on average KEYNESS values, regardless of significance of that figure. To further illustrate that point, refer to the values for the image macro keywords. They are the only keywords with an EFFECT value of 1 as they only function on KW and as such their appearance in the GerRefSr would be rather anomalous. They also represent the least number of tokens, which impacts their likelihood for a high overall KEYNESS.

5.2 Qualitative Analysis

What is clear when looking at the data from the quantitative analysis of KW, is that there is a significant lexical difference between KW and the larger German-language Reddit community as represented by the GerRefSr-Corpora. This difference is present in all four types of lexical items as can be seen in the above in table 5.1. In the following section, I will present and discuss the results of my qualitative analyses. First, I will focus on the keywords identified above, providing general results as well as highlighting specific items of interest, after which I will share the findings regarding my discourse analysis of KW user submissions.

A Qualitative Analysis of KW keywords

In order to discover what makes the keywords identified in section 5.1 unique, I employed the methodology laid out in subsection 4.4. This required the creation of five new categories to better organise my findings. First we have loaned lexical items (hereafter loans). These are lexical items 'borrowed' from one language to be used in another, in this case into German. Although his focus was placed on words, rather than my more broad lexical items, Hockett

(1958) offers a useful framework for categorisation, which I have adapted as needed, consisting of:

1. **Loan-items**, which are items adopted and integrated as a whole into the receiving language. For the sake of clarity, I will refer to these as **imports**.

2. **Loanshifts** represent words native to a language which have taken on meanings from another source language. These are also known as **semantic loans**.

3. **Calques** which are direct lexeme-for-lexeme, or morpheme-for-morpheme translations of a given item.

4. **Loan-blends** whereby a loaned element is combined with an element (or elements) from the receiving language.

The next category is designed to account for the lexical items related to 'bots', bots being automated programs designed to carry out various activities, such as posting, when triggered by a predetermined value or instance. The items counted in this category are only those whereby all instances attributable exclusively to bots. The third category is a catch-all for all non-loan lexical items of languages other than German. An example of an item from this catagory would be the Dutch word *kaas*, which occurs 112 times in these data, largely due to a raid[6] by members of the Dutch-language subreddit, r/Cirkeltrek. The fourth category, 'existent', accounts for all items with meanings native to the German language, including numeric items not impacted by loaning phenomena as described above. For the purposes of assessing whether

[6] A 'raid' is the name for targeted SPing on a given forum by members of a rival community.

a given meaning is 'existent', the Duden[7] was consulted. Usernames[8] make up the fifth category, regardless of whether or not they feature phenomena from other categories.

The results of my initial qualitative analysis are as follows[9]:

Category	# of Unique Items		# of Tokens	
ASCII Images & Characters	83	(8.20%)	7421	(11.115%)
Bot-activity Related	23	(2.30%)	974	(1.459%)
Existent	479	(47.90%)	12941	(19.383%)
Image Macros	54	(5.40%)	1534	(2.298%)
Loans	247	(24.70%)	11326	(16.964%)
Other Language	67	(6.70%)	798	(1.195%)
Subreddits	9	(0.90%)	30583	(45.808%)
Usernames	38	(3.80%)	1186	(1.776%)
total	1000	(100%)	66763	(100%)

Table 5.2: KW Keywords by type and quantity.

Two figures immediately stand out when looking at the above table. Firstly, the category Subreddits, despite making up less than 1% of all unique items represents nearly 46% of all tokens in these data. This is of course the result of the WSF-copypasta encountered during the quantitative analyses. Without the influence of this one item, subreddit names would make up only 0.129% of the total token count, a figure far more in line with the number of unique items. Again this serves to reinforce the impact SPing can have when pushed to its extreme, highly skewing quantitative data, to say nothing of is ability to disrupt discourse. The second figure of interest is the number of loans which account for just under 25% of all unique keywords. Despite making up only half as many keywords as existent German lexical items, there are nearly as

[7]The Duden's online version (available at www.Duden.de) was specifically used as it was the most accessible and, presumably, current, which would account for potential novel meanings not yet integrated into their print editions.

[8]While names of comment and submission posters were stripped from the corpora used to obtain keywords, usernames still appear if posted as part of a submission or comment.

[9]The percentages for the number of tokens have been rounded to the third decimal place and as such their sum is 99.998%

many loans-tokens in the data.

As stated above, there are four separate categories which I have used to group loan-types and better understand the breadth of the loaning phenomena present in these data. In the following subsection, I will present a general overview of these loans, by category, and share the items which I believe to most clearly exemplify the complex loaning KW members engaged in as they make meanings. Where applicable, an English translation will be provided in brackets directly after the lexical item in question. In the case of quotations, the username of the poster(s) will also be bolded for clarity, and line formatting condensed due to space-related considerations.

Imports

Given KW's general, albeit tongue-in-cheek, aversion to the English language, it is not surprising that there are few loan-items from English present in these data. In fact, there are only six such items (see table 5.3 below).

Arguably, *bannieren* and *fap* have their roots in languages other than English. Bannieren is attested to in Grimm & Grimm (1862), whereby it is listed as having been borrowed from French. The meaning whereby an individual rather than an act *per se* is banned is, for obvious reasons not present, and the verb itself is not listed in the Duden. I would argue that given the lexical closeness of the German verb bannen to its English cognate 'to ban'. KW users imported the item assigning to it morphological features typically reserved to predominantly French-origin verbs in an effort to disassociate its meaning from English. The other item, *fap*, and with it the derived verb *fappieren*, originate from Japanese manga (Tomberry, 2010), from which the onomatopoeia was transliterated first into English and then was spread via the internet. In that sense, it is only indirectly of English origin, however,

Lexical Item	Conjugations and/or Derivations	# tokens	Meaning(s)
bannieren	banniert, schattenbaniert, Bannierung	179	to ban, to shadow ban. A shadowban is when an individual can still post content, but only they are able to see it unless a moderator chooses to reveal it to the rest of the users.
Doxxern	N/A	14	Doxxers are individuals who access and publicise others' personal and typically private information, with the intent of enabling the harassment and/or shaming of said individual.
fap	fappieren	2517	'Fap' is an onomatopoeia for male masturbation; to fap is the verb.
Nigga	N/A	152	A slang term which is used to refer to another individual and is typical of African-American Vernacular English.
Pepes	N/A	8	A Pepe is an image of the anthropomorphic frog character from the cartoon Boys Club by Matt Furie. Pepes are typically used to express a diverse range of feelings such as anxiety, rage, sadness, and smugness.
Spammierungs-	N/A	82	A noun of the verb to spam. Due to how AntConc identifies tokens, the full item was recorded separately. The complete text is *Spammierungs- und Hochwählungs-pfosten.*

Table 5.3: The 6 imported lexical items among the top 1000 KW-keywords.

its use and spread can only be ascribed to it adoption into English (internet) slang. Further supporting this assertion are Urban Dictionary records for the item dating as far back as October 11, 2002, while earliest entries in German slang dictionaries are recorded no earlier than 2010 (*fappen* (Killerdildo)), and 2011 (*fappieren* (Anonymous)).

Loanshifts

Loanshifts account for a wide number of the loaned lexical items in these data. In total there are 37 such items. If you include derivations and conjugations (in the case of verbs) this number nearly doubles to 73. These loanshifts refer to a wide range of generally internet and pop-culture phenomena, as well as some Reddit-specific terminology. Let me first provide two examples of internet terminology which have undergone shift in German-language item meaning.

First we have the term *Nackenbart*. Much like the English term, neck beard, it originally denoted exactly that, a style of beard which covers a man's neck, without the inclusion of a moustache over the upper lip. In recent years neck beard has taken on a pejorative meaning as well, referring to slovenly, unkempt, typically overweight, and socially awkward men who spend excessive amounts of time on their computers, and the internet in particular. Such individuals are associated with pedantry, and an interest in obscure or taboo cultural material. They are also often associated with a highly over-enthusiastic consumption and festishisation of Japanese culture. The term neck beard is used not only to mock an individual's presumably unattractive outward appearance and character traits, but to simultaneously highlight their supposed lack of masculinity as well, as these individuals cannot grow a full beard as a more prototypical male can.

A second item which has seen a meaning-shift is the term *Soße*.[10] This item has taken on the meaning which *sauce* had previously acquired in many online contexts, whereby it almost invariably means *source*. Phrases including the term, or simply the item itself are now generally used to request or highlight information relating to the source of a meme, video, image, music, a statement,

[10]In the keyword list it appears as *sosse*.

etc, as can be seen in the excerpt below:

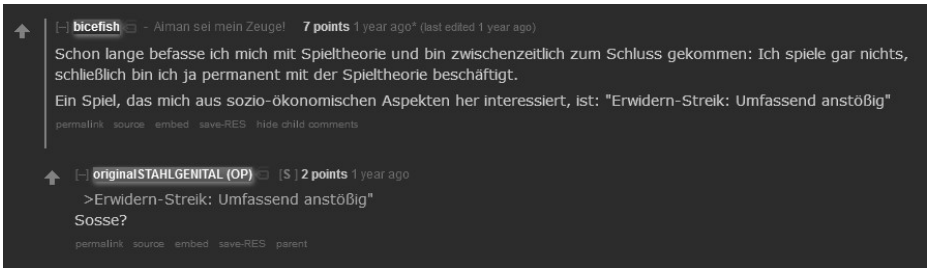


Figure 5.1: Use of the word *Sosse*, by the user *originalSTAHLGENITAL*.

Calques

The 65 calques (140 unique item tokens) represent the most numerous and diverse set of loaned items present in the keyword data, accounting for over half of all loaned lexical items. Present in these data are primarily compound nouns such as *DuRöhre* (YouTube), *Fledermausmann* (Batman), *Grüntextgeschichte* (green text story), and *Klickköder* (click bait). There are also initialisms of adverbial and adjectival nature, for example: *iel* (IRL - In Real Life); *LAL* (LOL - Laugh Out Loud); *MGW* (MFW - My Face When); and *NSFDA* (NSFW - Not Safe For Work). There are also number of calques relating to individual subreddits, some examples being */r/papawitze* (/r/dadjokes), */r/daspassierte* (/r/thathappened), and */r/ersteweltanarchisten* (/r/firstworldanarchists).

While majority of calques cannot be definitively attributed to members of KW - the first occurrences of a number of calques can be found elsewhere on the web, while the origins of most are entirely unclear - there are several which undoubtedly stem from KW users and have their roots in KW discourse. Of those which can be traced to KW, a number exemplify the apparent linguistic ingenuity of KW members and the diversity of processes which go into the

adapting linguistic items for their discourse. In the following subsections I will highlight several examples of these calques.

Treffauf

The first novel calque I will present is the noun *Treffauf* (meetup). In the overall context of Reddit, *meetups* are typically informal event for members of specific subreddits or redditors from specific regions to meet each other in person and get to know the members of their otherwise anonymous online community. Here we have a morphological calque whereby the 'creator' interpreted the *meet* in meetup to be the root of the verb *to meet* rather than equating the morpheme with the noun *meet*. This individual then applied the morpheme *auf* as the analogue to the English *up*. This novel item matches not only the morphemic structure of the English from which it has been derived but the syllabic structure as well: *treff* · *auf* - *meet* · *up*. While this may seem trivial, it shows a kind of linguistic consideration which other potential options would violate, like $*$*Kennenlernauf* for example. This was done despite the fact that the verb *kennenlernen* arguably better encapsulates the meaning of meet in this particular context.

Untentrepp

The next calque, the adverb *untentrepp* (downstairs) exhibits more of the considerations which played a role in the calque previously discussed. In this case we have have a novel lexical item which seeks to fill the same roles as its English counterpart while approximating its syllable structure as closely as possible given the available stock of potential morphemes. The user introducing the item, Klarname, even provides an implicit lesson as this novel calque fits the same contexts as the English item via the use of a so-called greentext

story:

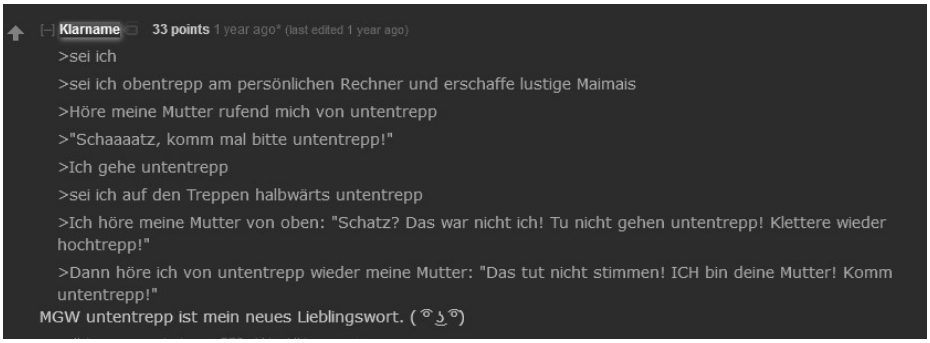


Figure 5.2: Use of the word *untentrepp*, by the user *Klarname*.

This story provides more than just the situational context surrounding *untentrepp*, it also reveals several other interesting linguistic novelties. Interestingly, the user also provides two further calques for the English adverb *upstairs* - *hochtrepp* and *obentrepp*. A probable explanation for this variation are the verbs which were used in those instances. It is far more common to use *oben* in conjunction with the verb *sein*, meanwhile *hoch* as an adverb indicates the direction of movement, and as such lends itself more easily to such verbs, in this case *klettern*. As such to only use the one or the other would potentially lead to confusion amongst the intended audience. However, a similar claim could be made for *unten*, as the morpheme *ab* more strongly indicates direction of travel.

Another linguistic feature which is visible in this excerpt, but which could not have been be identified by a method such as keyword analysis is the use of the Verlaufsform of *tun + infinitive*. This is a feature not typical of standard German, though it is found in several dialects. Such grammatical structures are more common of English (and other Germanic languages) as is the lack of the more strict verb-final placement for the infinitive, suggesting

a conscious effort to mimic English grammar. Another example of this would be the use of *rufend*, the present participle of *rufen*, to create a kind of ad-hoc progressive tense for the verb. This is because in its adjectival usage, *rufend*, can be translated as *calling*, which is a homograph and homophone for the verb *to call* when conjugated for the progressive tense, as it would be in the English translation of the same sentence: "Hear my mother *calling* me from downstairs". This is an instance of a semantic loan, whereby the existent item's meaning is expanded to include that from an outside source.

There are clearly many processes at play in the data above, exemplifying a number of different loaning phenomena, however there are others which highlight variation and creativity, and are show the inherent connection of KW to Reddit as a community and institution.

Lases, and KW Loans for 'Voting'

Given loaning phenomena which is so prominent within its discourse it is unsurprising that the central elements of Reddit, which plays host to KW and provides the framework upon which the community is built, should also be adapted within KW discourse. The most opaque result of these phenomena, in terms of deciphering its meaning from the position of an outsider would be the KW calque for the name 'Reddit' itself - *Lases*. When broken down into its constituent lexemes *las* and *es* the meaning becomes more clear. The first of these, *las* is the simple past tense[11] of the verb *to read*. The connection comes from the initial syllable of Reddit, which is a homophone for *read* the simple past tense of *to read*. This is an intentional play on words by the site's creators, and the calque reflects KW member's knowledge of this.[12]

[11] Indicative and third person singular forms.

[12] Without this knowledge one could potentially assume that the name derived from the word *red* and thus result in a the calque ***Rot****es*.

With *Las* clarified, the addition of the lexeme *es* is a relatively obvious one. However, what may potentially throw a reader's ability to understand this calque is the way in which the syllable boundaries shift when these lexemes are combined. This is because in German phonology, the consonant [s] can function as syllable joint,[13] whereby it acts as the coda of a preceding stressed syllable, in this case *la*, and the onset of the following unstressed syllable, *es*, effectively joining them together (Onysko, 2007, p. 231). This blending of syllables can obfuscate the lexemes involved, potentially leading a reader to interpret the item as *lase-s*, especially given that the, albeit obscure and archaic, noun *Lase* (beer barrel) exists in German.

As a lexical item, *lases* has been fully integrated into KW-discourse. There is a diverse range of compound nouns featuring the item including: *Lasesgold* (Reddit gold); *Lases-Diener* (Reddit servers[14]); *Lügenlases*[15] (lying Reddit[16]); and *Unterlases* (subreddit).

Just as *lases* is result of borrowing processes, so too are the various items relating to the up- and downvoting voting system of Reddit. Again these items have been integrated in to the KW vocabulary as nouns, and (separable prefix) verbs. Beyond that there is a moderate amount of over-lexicalisation which is demonstrated by these items. Take for example the three calques for the verb *to upvote* and its noun, upvote: *aufwählen*, *Aufwähl*[17]; *raufwählen*, *Raufwähl*; and *hochwählen*, *Hochwähl*. Of the three items *hochwählen* appears to be the most widely accepted, with 14 lexical items present in the Keyword list compared with just two for *aufwählen*, and five for *raufwählen*. It also

[13]This phenomenon is not limited solely to the consonant [s].

[14]Servers here refers to computer servers, adding an element of wordplay in the form of a loanshift of the meaning of *Diener*.

[15]This is a play on the term *Lügenpresse* which has seen a resurgence in recent years due to its use as a pejorative chant by members of the PeGiDa movement.

[16]More literal translations would be *lies Reddit* or *Reddit of lies*.

[17]Present in the keyword list as the plural *Aufwähls*.

indicates the number of possibilities which KW members have determined to be reasonable analogues for the English morpheme *up*. The calque of downvote, demonstrates less variability, as only *runtergewählt,*[18] its substantives *Runterwähl* and *Runterwähls*, and the slightly less colloquially expressed form *Herunterwähl* are attested in the keyword list. Interestingly enough, this has parallels to the calque *untentrepp* discussed above, as the potential alternative *ab* was eschewed in favour of inflections of *unten*, while selections for *up* saw a greater level of diversity.

There are 54 other calques present in the keyword data; however, it is beyond the scope of the present work to provide a thorough breakdown of the processes behind their 'creation'. With that in mind, I would like to now turn my attention to the final category of loaned-items, loan-blends.

Loan-blends

With only 16 items recorded within the keyword list (25 when taking variations into account), loan-blends make up the second smallest group of loaned lexical items in those data, just ahead of imports. These items are largely adaptations to English terms made in order to align their orthography more closely with German phonological conventions. Examples of this would be *Opi* [opiː] (cf. *OP* [opʰeː]); *Miem* [miːm] (cf. *meme* [meːmə]); and *Gugel* [guːg əl] (cf. *Google* [goːglə]). There are also instances of existing lexical items being adapted to match usage conventions of their English-language counterparts such *Ruhe in Frieden* which is modified into *RIF in Frieden*[19] (RIP in peace), and *MFW*. The lexical item *MFW* conforms the initialism, which would otherwise be

[18]The past-participle of *runterwählen* which is in the KW-Corpora, but not the keyword list.

[19]Only *RIF* is attested to in the keyword list, however, in the KW-Corpora, 60 out of the 86 instances of this item are as part of the phrase *RIF in Frieden*, a loan of the English-language meme.

loaned as *MGW*,[20] to the spelling of the English initialism *MFW* which means 'My Face When'.[21] This is done by substituting the lexeme *Gesicht* with *Fresse* which carries the same meaning when used colloquially.[22]

Given the variety and breadth of loan items in the used in KW-discourse, two things strike me as particularly apparent. Firstly, there is a high level of linguistic creativity and knowledge, whether conscious or otherwise, which plays a central role in the discourse. There are novel calques which rely on the skilled manipulation of morphemes and lexemes, as well as loan-shifts which make reference to the items and tenses which are translated into German while simultaneous shifting an existent items meaning to that of the donor language. There are also the loan-blends presented above which manipulate spellings with conscious attention to orthography so as to leave the phonology of items unaffected by the loaning process. Secondly, for the vast majority of these loans to work, one must have a very well-rounded knowledge of English grammar, English-language Reddit and internet culture, as well as English-language pop culture more generally. In fact, I would argue that presented loans, which form a core element KW-discourse, demonstrate that a comprehensive knowledge of English is as critical as German linguistic and cultural knowledge in order for one to effectively integrate into, and participate in, the KW-community.

As of now the focus of my analyses have been the oft mentioned lexical items which form the foundations of KW-discourse, but aside from a few brief excursions, little explicit attention has be paid to the discourse itself. As stated above, such an analysis is crucial to obtaining a more balanced understanding

[20] *Meine Gesicht Wenn* (My face when).

[21] This initialism is used to emphasise an individuals reaction to a situation or event, and is typically used in conjunction with an image of a face representing said reaction.

[22] *Fresse* is typically used pejoratively due to it being derived from *fressen* which means *to eat (like an animal)*.

of any given discourse community, such as KW. While a truly comprehensive analysis goes beyond the intended scope of this thesis, in the following section I will provide the findings of my initial investigations of KW-discourse as they relate to user submissions. Afterwards, I will discuss whether or not the Hallidayan concept of anti-language can be applied to the discourse recorded in these data, and if so, to what extent.

5.3 An SFL-based Discourse Analysis Focused on the Context of User Submissions and their Reception

As detailed in chapter 4, subsection 4.4, a total of 55 user submissions were analysed using systems of features relating to register as conceptualised within SFL. These submissions were divided into five distinct groups prior to analysis based on their statistics relating to the submitting user, their percentage of upvotes and total upvotes, and a control group of randomly selected submissions. In keeping with these groupings, the results of my analysis are presented below in subsections dedicated to each group. Afterwards, I will summarise my findings as they pertain to how KW reacts to the register (i.e. context) of user submissions. However, prior to these group-specific and generalised observations, I will first discuss the systems and their features which were either identical across all submission groupings or were otherwise indeterminable. This will be followed with a brief discussion of my rationale for determining certain features within the TENOR subsystem of SOCIAL DISTANCE.

The Standard & Indeterminable (sub)Systems and Features of all Submissions

Given the nature of KW as an online, semi-anonymous discourse community which is in many ways restricted to a specific subdivision of a specific wesbite, there are certain contextual variables which remain the same for all the interactions, in this case submissions, which can be analysed. The first of these which I will address is the contextual variable of MODE and its subsystems.

Looking first at the ROLE OF LANGUAGE subsystem, we cannot escape the fact that it is constitutive of the activity of posting a submission. Reddit will not allow a submissition to be posted without a minimum of a single character. While it may seems that a submission of a single character may carry no significant interpretable meaning, I would argue against that. In fact there is a submission (see subsection 5.3) in these data which supports this assertion. Even simply supplying a URL as a submission title conveys meanings and is, in this sense, a linguistic act. For those reasons ROLE OF LANGUAGE remains the same across all data. This consistency also holds true for the CHANNEL and MEDIUM subsystems. In the case of CHANNEL, it is restricted to WRITTEN as far as what the users themselves are supplying to the submission. One might argue that submissions linking to images or other media impact CHANNEL, and to an extent I would have to agree. However, these additions, while certainly important, are external to what the users themselves are supplying to a given submission. In that sense they are supportive of, and secondary to, the written communication which is present. For that reason WRITTEN takes precedence when accounting for the CHANNEL subsystem. MEDIUM exhibits similar characteristics in that it will always be internet-facilitated communication via Reddit.

Moving on to the subsystems of FIELD, we encounter similar consistency

within the MATERIAL ACTION subsystem as we have in the subsystems previously discussed. In the contexts of a submission the only feature to be selected is OBLIGATORY. This is because users must engage with an external, be it with a computer or smart phone, interface in order to participate in the act of submitting a submission and therefore communicate to the KW-community. Next we have ACTION WITH SYMBOLS which will always be NECESSARY given the format of a Reddit submission. The only instances whereby Reddit-based meaning making can satisfy the UNNECESSARY feature, would be the use of up- or downvote buttons, which convey a reaction to language without using language itself. With that in mind ACTION WITH SYMBOLS is consistent up to the feature POST at which point the system grows in terms of potential selections. SPHERE OF ACTION exhibits a similar phenomenon, although far more levels of delicacy are involved.

All user submissions on KW consist of the following string of SPHERE OF ACTION subsystem features:

QUOTIDIAN → INSTITUTIONAL → PRACTICE → GENERAL → ROUTINE

This is because Reddit submissions on behalf of the average user, i.e. those engaged in the discourse which I am analysing, are a routine practice linked to the institution of Reddit; this institution determines the available forms of communication as well as dictating to a certain extent the limits of their content. They are also quotidian in that engaging in Reddit-based discourses requires no long-term commitment (Wegener, 2011) or specialised knowledge beyond knowing how to use a web-browser.

The GOAL ORIENTATION subsystem of the FIELD network has been excluded from analysis. Lacking in explicit data directly obtained from the users themselves, I felt it inappropriate to attempt to assign a particular goal to their submission and the features thereof. This is because such an as-

sessment relies on a truly subjective interpretation of a submission. Such an assessment would necessarily be influenced by its actual results, rather than intended outcome, as I have no data which provides any indication of a user's intentions.

All four subsytems of TENOR have features which are unchanging with regards to submissions. SOCIAL DISTANCE with regards to the possible options for the CODALLY DISTINCT feature are always STRONG CLASSIFICATION and STRONG FRAMING. This is due to distinct and clearly identifiable roles played by both submitter and commenter. They are inherently obvious given the interface presented to users and visitors of Reddit, regardless of the subreddit. Of the AGENTIVE ROLE, it is both ACQUIRED and RECIPROCATING → COMPLEMENTARY in nature. Access to the role of a submitting user is acquired when signing up for an account with Reddit, and is COMPLEMENTARY due to implicit invitation for feedback inherent to the nature of Reddit-based discourses. Looking at the SOCIAL HIERARCHY of submissions, it consists of the following feature selections: NON-HIERARCHIC → CHOSEN → PEER GROUP RECREATIONAL; DECLARED; and IMMUTABLE

There is no specific hierarchy tied to posting a submission to Reddit, beyond perhaps those tied to privileges given to moderators, none of which were exercised in submissions examined. The hierarchy is also explicit in that the role of submitter is clearly identifiable via the user interface, and immutable as the submitter always remains identified as such within a given instance. The subsystem of NETWORK MORPHOLOGY contains both SCALAR and NON-SCALE FEATURES. The non-scalar features closely mirror those of AGENTIVE ROLE, in that they are DYADIC COMPLEMENTARY ROLES split between submitter and reader, and therefore POSITIONALLY DEFINED ROLES as well. Looking at the scalar features, DENSITY can only be determined relative to other subreddit

discourse-communities, in which case it is LOW at only approximately 3,300 members. Exact DIVERSITY is impossible to determine without reliable demographic data, which I do not have access to. DIRECTION is BOTH WAYS as users have the opportunity to provide a visual evaluation of a submission via the voting system, as well as the opportunity to post comments in reply. The CENTRALITY of the observed linguistic acts is also HIGH given that the submission serves as the focal point for further discussion, and these are further centralised within the KW-subreddit. With that in mind CLUSTERING is also HIGH.

At this stage I would like to return to the SOCIAL DISTANCE subsystem as it contain a series of features which cannot be determined given the data available to me. These features deal with the number of connections between individuals, that is, if their relationships are MULTIPLEX or UNIPLEX and the contexts surrounding them. Given the semi-anonymous nature of Reddit, it is impossible for me to determine whether or not any given users have any relation to the other outside of the site. For that reason I have not included it when mapping the systems of the individual submissions. With regards to how I interpreted the features which follow from CODAL SHARING. LOCAL HISTORY IN COMMON was determined to be present for any submission posted by a user with prior posts on KW. CULTURAL CAPITAL IN COMMON is of course difficult to determine when analysing data from semi-anonymous sources, as it is impossible to determine what culture an individual is most familiar with. Given that the interactions are via Reddit, it can be assumed that there is a baseline of commonality as it relates to the norms and culture of the site. There is, however, one option available to me which I made use of when analysing these data, and that is the language of the submission. I have chosen to identify a submission as having the feature CULTURAL CAPITAL

IN COMMON if the post is written in German (regardless of dialect). This is because language is the only strongly identifiable cultural element immediately available to the reader of a submission, and the use of German is one of the foundations upon which KW is explicitly built.

With the above systems and features duly explained I will now move on to the results of my analysis and a discussion thereof. In the following subsections I will briefly outline data related to each group[23] followed by my findings as they relate to their contextual variables.

Submissions from One-Time Posters to KW with 100% Approval

This group consists of 11 submissions and range in karma from 6 to 70 points (Avg. 24; median: 17). Looking at the FIELD subsystem of SPHERE OF ACTION starting after the common feature of ROUTINE (see subsection 5.3), 8 out of 11 submissions contained the feature ESTABLISHED, after which the feature MEME was selected. It is at this point where the selected features become more diverse. After selecting for MEME, nearly all submissions universally extended one further layer of delicacy beyond this feature, though they generally did not exhibit multiple selections at this layer; only two submissions contained more than a single MEME feature, with two and three features being selected in those cases. There was also only a single submission in this grouping which contained MEME without extending to a further level of delicacy. No other features branching off from ESTABLISHED were present in these data. Contrasting with the submisions which selected for ESTABLISHED, are three which instead contained the feature NOVEL. Of those, all three went on to contain the feature NEW TOPIC.

[23]See chapter 4, subsection 4.4 for a description as to how these groups were selected and organised.

Moving on to the FIELD subsystem of ACTION WITH SYMBOLS, all submissions go on to select for OC SINGLE USE, at which point six then contain NON-PARTICIPATORY while the remaining five contain PARTICIPATORY. Interestingly, all three submissions with the SPHERE OF ACTION feature NEW TOPIC are among those with the feature NON-PARTICIPATORY. There is, however, no data to suggest any sort of relationship between these two selections. All other possible features extending one layer of delicacy beyond POST have been selected at least once in these data; the feature VIDEO is present four times, while IMAGE occurs only twice. TROLL appears just once in these data whereby the features at next layer of delicacy are OTHER ORIENTED and TEASING.

The features of the SOCIAL DISTANCE subsystem of TENOR present in these submissions are entirely identical. In every case, the possibly variant features are inevitably INCIDENTAL → CHANCE, given that these are posters users with no prior record on KW, thus assuring the CODAL SHARING feature of NO LOCAL HISTORY IN COMMON as well. All of these submissions also share the feature of CULTURAL CAPITAL IN COMMON, as the language present in each is German.

Submissions from One-Time Posters to KW with Low Approval Ratings

With only seven submissions fitting this criteria, this represents the smallest of all groups considered in this portion of analysis of KW and its discourse. The approval ratings for this group range from 58 to 67% (Avg. 64.44%; median: 65%), with karma scores between 4 and 7 points (Avg. 5; median: 5), which are quite low relative to all other groups, with the exception of the zero-score group (see the next subsection below). When analysing the SPHERE OF ACTION subsystem, the first feature to stand out does so, not because of

its inclusion, but rather its absence; not one submission in this group contains the feature MEME, which is in immediate contrast to the group discussed above in the preceding subsection. Instead, my analysis uncovered the features NATIONALITY (2 GERMAN; 1 AMERICAN), ETHNICITY (1 instance), and REGION (2 instances of COUNTRY → GERMANY, one of which contains the feature LANDKREIS), when a submission first features the prerequisite ESTABLISHED. It should also be noted in all six of these submissions, only one feature at the delicacy layer beyond ESTABLISHED is present, rather than potentially more complex constellations of features. Of these seven submissions, only one contains the feature NOVEL, whereby NEW TOPIC alone is also present.

When looking at the ACTION WITH SYMBOLS subsystem, six submissions contain the feature TROLL, at which point the further levels of delicacy selected are somewhat evenly represented: 2 CONTROVERSIAL; 3 RAID → FRIENDLY; 1 PERSONAL ATTACK → TEASING. The single post which does not contain TROLL is also the only submission whereby the features OC → SINGLE USE → NON-PARTICIPATORY are not present as well. This is due to the submission containing the features REPOST → EXACT which precludes the aforementioned possibility of OC. The majority of these submissions (4) also contain the feature IMAGE, while VIDEO is present only once. There are just two submissions whereby the features SELF → SUBREDDIT APPROPRIATE of this subsystem are present.

As with the submissions from one-time posters with 100% approval discussed above, all submissions of this group contain the features INCIDENTAL → CHANCE, and thus the CODAL SHARING feature NO LOCAL HISTORY IN COMMON, when analysing the TENOR subsystem of SOCIAL DISTANCE. However, these data show some variance when it comes to the CODAL SHARING, as there is a near even split in when looking at CULTURAL CAPITAL IN COMMON.

Four submissions contain the feature, while three present NO CULTURAL CAPITAL IN COMMON instead. These three submissions are the same submissions with the RAID → FRIENDLY features. They stem from a raid by members of r/cirkeltrek, a Dutch-language CJ-subreddit along similar lines as KW.

The Top Submissions of 2015 on KW According to Karma Scores

For this group of 12 submissions, the karma scores range from 239 to 472 (Avg. 315; median: 290), while approval ratings range from 89% to 99% (Avg. 95.17%; median: 95.5%). These represent the most highly upvoted submissions on KW, and have by far the largest karma scores of all submissions analysed; the lowest score is still some 74 points greater than the highest score (165) found amongst the other groupings of submissions. Looking at the features present in these submissions, it is easy to see what is likely behind this; all 12 present the chain of features ESTABLISHED → MEME → BEHAVIOUR → VOTING → UP, from the SPHERE OF ACTION subsystem alongside feature ENCOURAGING from the same.

No other ESTABLISHED, or NOVEL features are present in the SPHERE OF ACTION subsystem of these submissions.

In the ACTION WITH SYMBOLS subsystem, 11 out of 12 submissions contain the features OC → SINGLE USE → PARTICIPATORY, while a single submission instead presents the features REPOST → MODIFIED. Furthermore, 11 of 12 submissions, including that with the feature REPOST, also possess the features TEXT → COMPLEMENTARY → IMAGE, whereas only one has features TEXT → SELF.

Moving from variable of FIELD to that of TENOR, a similar pattern is repeated as again 11 of 12 share identical features of a single subsystem, in this

case SOCIAL DISTANCE. These shared features are REGULAR → RECURRENT alongside the CODAL SHARING feature LOCAL HISTORY IN COMMON. Even though one submission does not have the SOCIAL DISTANCE of LOCAL HISTORY IN COMMON, all 12 do exhibit CULTURAL CAPITAL IN COMMON.

Zero-Karma Submissions

These 12 submissions all have the dubious honour of having a zero net karma score - the absolute minimum displayable score possible for a submission on Reddit. While their karma scores appear the same, we can still view their relative approval based on their upvote percentages. The range of approval present in these data is from 22% to 50% upvotes (Avg. 36.8%; median: 38%).

These submissions show a wide range of features within the SPHERE ACTION and ACTION WITH SYMBOLS subsystems of FIELD, just as they do in terms of features of the SOCIAL DISTANCE subsystem of TENOR. Starting with SPHERE OF ACTION there are only three submissions with the feature MEME. Up to this point MEME has solely been a feature found in positively received submissions, whether they be from one-time or recurrent posters. With that in mind it makes little sense to include them in a generalised accounting of this group, as I have with the previous collections of submissions.

Breaking down the first of these submissions, we quickly encounter the explanation for one of these instances which is revealed through the string of MEME features which are present, namely MEME → ENCOURAGING + BEHAVIOUR → VOTING → DOWN. This indicates that the user was quite literally asking for downvotes, and in a round-a-bout way almost indicates a level of approval for this submission. This acceptance, if paradoxically exhibited, would also be expected given that the SOCIAL DISTANCE subsystem features of LOCAL HISTORY IN COMMON and CULTURAL CAPITAL IN COMMON are also

present as they have been in the majority of positively evaluated submissions.

What then of the other two submissions with the feature MEME? In one instance we have the first, and in these data only, submission whereby the variable MODE plays a role. This is due to the fact that the user attempted to integrate an image link into their submission. Unfortunately, they failed to format the link correctly which prevents it from opening when accessed by other members of KW. This effectively violates the given norms relating to the CHANNEL subsystem as it is *de facto* unable to be fully realised. Ultimately, the consequence of this error is a collapse of the coherence within the submission, which in turn causes its rejection by the others. This also leads the submission to possess the SHITPOST feature of the ACTION WITH SYMBOLS subsytem.

The final negatively appraised submission with the feature MEME, also possess a feature which has only been encountered in low and negatively received submissions, the SOCIAL DISTANCE subsystem feature of NO CULTURAL CAPITAL IN COMMON. While not immediately apparent in the submission title, the supplied image contains a meme of a famous American actor along with English-language text which in turn assigns the CODAL SHARING feature NO COMMON CULTURAL CAPITAL which is common amongst the low and negative approval submissions.

The remaining nine submissions belonging to this group display a range of features which have been previously associated with the poor reception of submissions. Two thirds contain the feature TROLL, and of the remaining three submissions not yet touched upon all contain the feature SHITPOST, albeit in two different configurations: (2) SHITPOST → COPYPASTA + OBSTRUCTIVE; and SHITPOST → LOW EFFORT. Looking at the TENOR features which appear in the nine submissions, which I have just highlighted, we come across the rather interesting division. There are four submissions with the features NO

LOCAL HISTORY IN COMMON + NO CULTURAL CAPITAL IN COMMON and four submissions with the polar opposite arrangement of features, i.e. LOCAL HISTORY IN COMMON + CULTURAL CAPITAL IN COMMON. Of the former group, all contain the ACTION WITH SYMBOLS feature TROLL, while later consists of one submission with that feature as well as three more with the feature SHITPOST from that same subsystem. That these submissions were so negatively received despite sharing a constellation of features which was previously associated with highly approved posts suggests that for KW discourse the ACTION WITH SYMBOLS subsystem of FIELD carries more weight with regards to group acceptance than the SOCIAL DISTANCE subsystem of tenor. The final submission features mixed CODAL SHARING features whereby there NO LOCAL HISTORY IN COMMON and CULTURAL CAPITAL IN COMMON are present. The SPHERE OF ACTION subsystem feature ETHNICITY inclusive of its systemic prerequisites is present, as is the TROLL → CONTROVERSIAL features from ACTION WITH SYMBOLS.

The Control Group of Randomly Selected Submissions

The submissions described here were selected using the process laid out in chapter 4, subsection 4.4. The purpose of this group was to collect a group of what could be considered 'average' KW submissions in order to document the typical features of KW-discourse while mitigating an potential selection bias on my behalf. In keeping with the groups of zero-karma, and top-scoring submissions, there are 12 submissions in these data. Their karma scores range from 31 to 165 (Avg. 62; median: 46.5), while their approval ratings range from 88% to 100% (Avg. 94.3%; median: 94.5%).

Of all the SPHERE OF ACTION features possible, these submissions possess primarily MEME. Of these, the majority (8) exhibit the features BEHAVIOUR →

ENCOURAGING + VOTING → UP, while one does not extend in delicacy beyond the level of BEHAVIOUR; several (6) of these submissions also possess further MEME features, for example: (2) INSTITUTION; and SPECIFIC USER → OTHER. Keeping within the same subsystem, one of the remaining four submissions contain the features MEME → ITEM OF INTEREST + PUBLIC FIGURE, while three diverge from all other submissions after the POST layer; all then exhibit the same set of features NOVEL → NEW TOPIC.

When taking ACTIONS WITH SYMBOLS the submissions in these data tend to feature more complex constellations of features, which strikes me a logical given their selection at random; all 12 do contain the feature OC with the following added layers of delicacy: (1) EXPLOITABLE; (8) SINGLE USE → PARTICIPATORY; and (3) SINGLE USE → NON-PARTICIPATORY. At the same level of delicacy as OC, five submissions contained the features TEXT → SELF, whereas the other seven contained (6) IMAGE; and (1) LINK TO WEBPAGE INTERNAL, all of which follow from the features TEXT → COMPLEMENTARY. One of these submissions also exhibited SHITPOST in addition to its other features. This submission is particularly interesting because it consisted entirely of a single letter, an 'F'. The assumption which I, and apparently the 25 KW members who commented on it, made was that this post relates to the 'press F to pay respects' meme. This particular meme sprung up out of a particularly absurd moment in the game Call of Duty: Advanced Warfare, whereby a supposedly solemn moment at funeral for the player character's best friend and comrade is interrupted by a small prompt for the player to press the F key to pay their respects (Morris & Brad, 2014).

With regards to TENOR, the overwhelming majority of the submissions (11) in this group possess the CODAL SHARING features LOCAL HISTORY IN COMMON + CULTURAL CAPITAL IN COMMON, as would be expected for 'av-

erage' exchanges amongst members of a relatively small discourse community such as KW. There is, however, a single submission which exhibited the feature NO LOCAL HISTORY IN COMMON in addition to CULTURAL CAPITAL IN COMMON.

At this point I would like to summarise and discuss the above findings so as to present a cohesive overview of which contextual variables (i.e. registers) appear to play key roles in influencing KW-members reactions to user submissions. I will also reflect on my findings in relation to the keywords presented above in section 5.1 and subsection 5.2.

5.4 KW Keywords and the Community's Reaction to Register

When contrasting the positively received submissions against those which were uniformly or largely received negative appraisals by KW members, several clear trends become apparent. Firstly, and most importantly, the TENOR of the attempted communication must possess the CODAL SHARING feature CULTURAL CAPITAL IN COMMON. At a bare minimum this means that a submission must be in the German language, though those which also include KW keywords also tended to be received more positively as determined by average karma score and upvote percentage, even when excluding the top voted submissions, as can be seen in the table below:

Category	# of Submissions	Avg. Karma	Avg. Upvote %
w/o keywords	16	21	89.2%
w/ keywords	23	194	95.7%
w/ keywords excl. Top Posts	11	63	96.3%

Table 5.4: A breakdown of the submission scores based on the inclusion of lexical items from the list of the top 1000 Keywords

Interestingly the feature whether or not an individual has any apparent prior connection to KW does not appear to play a major role in terms of approval as a percentage of upvotes as long as CULTURAL CAPITAL IN COMMON is present. The difference between the two poles of the feature (LOCAL HISTORY IN COMMON) is associated with a variance of only .01%, and indeed in favour of NO LOCAL HISTORY IN COMMON (86.04% vs. 80.05%). However, this relationship reverses somewhat to 94.7% vs. 90.53% when top and zero-karma score submissions are removed from consideration. A submission by an individual with local history in the KW-community is associated which significantly higher karma scores regardless of whether the top and bottom performing submissions are considered. What this means is an individual's submission will likely receive more attention if they are already known to other members of the KW. With that said, regardless of whether or not a given poster is familiar to KW, submissions are very nearly as likely to receive a positive evaluation from the members who do engage with it.

The apparent conventions surrounding FIELD and its subsystems of SPHERE OF ACTION and ACTION WITH SYMBOLS are necessarily more complex, given the vast array of possible topics of discussion. However, what is clear in these data is the importance of memes both as a systemic feature and more broadly within KW-discourse. Nearly every positively evaluated submission, as evidenced by their karma scores and upvote percentages contains, this feature; it is the rule rather than the exception that successful attempts at joining in the discourse of KW make reference of their memes or lexicon. These are as integral as the German which forms the foundation more broadly speaking.

So what then are the features which typically meet with disapproval? Looking at the data, the most consistent are the ACTION WITH SYMBOLS features TROLL and SHITPOST which are met near universally with downvotes

regardless of which features follow from them. This is may seem surprising given their general association to CJ-discourse. From an outsider perspective, these communities are largely places where members ceaselessly SP and troll each other. However, such a viewpoint observes the discourse from the wrong perspective. I argue that these are communities in which members can engage in otherwise undesirable discourse without being entirely disruptive to discourses which they may otherwise take more seriously. The key distinction to keep in mind is that these acts are done in an ironic fashion for their satirical value. They are a way for members of these communities to mock, critique, and violate the norms of a larger community which they themselves are a part of. However, that also necessitates the creation of their own set of discoursal norms which can themselves be violated. Thus not all trolling and SPing is created equally, and without a clear understanding of a given community, an SP can drift from desirable and accepted to out-of-place and rejected, just as those in the analysed submissions have.

Now that I have presented the findings of my analyses, I will discuss KW in the context of anti-language, and whether or not there are sufficient data to make that connection.

KW and Irony

One element the KW-discourse which certainly deserves mention is its highly ironic nature, not merely in terms of the lexicon used but also the overarching narratives and subject matter. Before highlighting specific instances of irony in these data, I will first provide a brief definition of what is meant by the term. Contrary to what may be the more widespread interpretation (Hutcheon, 1992) of irony, what is meant here is not antiphrasic substitution of a one meaning for another, i.e. saying one thing but meaning something

else which is essentially the opposite. Rather, irony is "the interaction not only between ironist and interpreter but between different meanings, where both the said and the unsaid must play off against each other (and with some critical edge) in order for such a process even to be recognized as ironic... a mixture of the pragmatic (in semiotic terms) and the semantic, where the semantic space is a space 'in between,' comprising both the spoken and the unspoken"(Hutcheon, 1992, p. 220).

The above definition is clearly more complex but more appropriately encapsulates the highly nuanced nature of irony. This nuance can be observed in KW-discourse at all levels. Let us first look at the content of a user submission:

Figure 5.3: A user submission regarding the difficult relationship between GEMA and YouTube.

Here we have a user, CryEagle, make reference to the often difficult business relationship between German music and artists' rights management organisation GEMA and the online video content host YouTube. The irony of the submission stems not solely from the content written by CryEagle, i.e. "1 Hochwähl = 1 Einigung (1 Upvote = 1 Agreement)", but also the interplay between the user's message and the message provided by YouTube with regards to why the content, although available to users elsewhere, is not accessible for

users within Germany. There is also the 'critical edge' which underlines the unspoken but clear frustration of a German internet user who is not able to access German content, while those abroad have unrestricted access. There is also irony present in the use of a an imported term from English - the calque *Hochwähl* (upvote) - to reference a desire for access to German cultural content presented to the world by an American corporation but made unavailable to Germans by a German corporation, whose mandate is to protect and promote German cultural goods.[24]

The ironic elements of CryEagle's submission are further underlined by comments posted in reply:

Figure 5.4: The two top replies to CryEagle's submission.

In both comments we can see irony clearly expressed in the discourse; there are intentional overstatements of the seriousness/extent of the frustrations felt by internet users both towards towards GEMA - GEMA gehört ausradiert weil wortwörtlich Satan, binichrechts? (GEMA should be exterminated because [it is] literally Satan, amirite?) - and in the application of slogans relating to controversy and scandal, namely through the use of *#GEMAgate*[25] and *Es ist*

[24]There is a great deal of debate about the methods, reasoning, and effect of GEMA's (internet) rights-management. Regardless of where one might stand on the issue, what is clear is the frustration on the behalf of all parties that the situation has yet to be resolved.

[25]In recent years man scandals, particularly political scandals have been referenced using *-gate* as a suffix. This usage of *gate* arose out of the Water*gate* scandal involving former U.S. president Richard Nixon which erupted in the early 1970s. The use of # is also a nod to so-called 'hashtag activism', that is the use of social media to spread political messaging, which is often punctuated by the use of # (a hash mark) to tag key words.

deswegen, dass ich den Feminismus benötige (that's why I need feminism).[26] Just as the above definition of irony suggests, the reader of user postings require more than a simple assumption that the users mean the opposite of what they are writing, but rather a detailed knowledge of pop-culture, current and historical affairs, and an ability to navigate the multiple and compounding[27] meanings being presented.

Due to the importance of interaction between 'ironist and interpreter', to use Hutcheon's terms, the necessity of a shared history between participants, or at a minimum an understanding of the Contexts of the participants' discourses is clear. This further underlines the separation between KW and the greater German-language Reddit community. A particular awareness of lexical and discoursal norms (such as extensive use of irony) is required to engage with KW as its members do. In the following section, I will continue to touch on this aspect of KW-discourse, as I examine KW within the framework of anti-language.

5.5 KW and Anti-language

The first element to consider when determining if KW as a discourse community makes use of an anti-language in their discourse, is whether or not the members themselves constitute an anti-society (see chapter 3, subsection 3.1). This would, of course, require that the individuals active within the KW community engage in behaviour which is largely relegated to the fringes of the greater community of which they are a part. I argue that given the great value placed on CJ-style discourse and community-specific lexical items directly con-

[26]This was a phrase used widely in online (primarily) social media campaigns to spread feminist messaging.

[27]All aspects of the comments build their meanings upon the framework provided by the initial submission.

flicts with, and offers a conscious alternative to, the discourses found in the subreddit communities which are represented within the GerRefSr-Corpora. The kinds of submissions which dominate KW in terms of group approval are clearly out-of-sync with those of the GerRefSr-communities in that they are intensely meme-focused. r/de, for example, features such a breadth of discourse that the subreddit offers methods for sorting submissions based on their general theme (see chapter 2, subsection 2.2). None of the memes encountered in process of my discourse analysis of KW-submissions could be traced to any of the subreddits included in the GerRefSr, including the CJ-oriented subreddit r/SCHLAND, indicating that their discourses are also substantially different in that regard. What this equates to is significantly different set of discoursal norms between the communities considered, which in turn indicates a separation between their societies.

In becoming members of a separate subreddit it is a clear choice on behalf of the KW discourse-community to signify their bond to said community. This is the case even if their bond is not immediately visible until their actions within the community are evidenced by some sort of contribution. This is consistent with the members of anti-societies as discussed in Halliday (1976), whereby the membership within these societies is typically undeclared or known except to members of the same; no given Reddit member would be able to identify a KWer by username alone unless they too ran in the same circles, so to speak, or the KWer's identity was explicitly exposed as such. Indeed, in this regard there are many parallels to the criminal anti-societies highlighted by Halliday. In both contexts, membership is held only by a small subset of the larger community, and generally kept to oneself given the niche and typically undesirable nature of the activities in which they engage.

There are also parallels to other groups analysed in the framework of anti-

society, whereby said groups have chosen to separate themselves from the larger society in order to 'preserve' what they feel defines them (cf. Schniedewind, 1999; Domeris, 1999). In the case of KW, this element which they are seeking to protect is the use of the German language. Their method of achieving this is by eschewing the use of English-language lexical items which have been imported into the German internet lexicon but whose lexemes and morphemes remain identifiably English to an otherwise untrained eye. In order pursue this agenda, as satirically motivated as it may be, English Reddit and web-terminology is borrowed and consciously manipulated using a number of linguistic methods, and thus appropriated by KW-members. One need not necessarily analyse KW in order to uncover this protectionist streak underlining much of their vocabulary; it is plainly laid out in the sidebar of every page of the subreddit. However, to use a rather ironic proverb given the context, actions speak louder than words, and empirical evidence that this activity has truly been adopted by the community weighs more heavily than a simple declaration put in place by an unidentifiable moderator.

Looking directly at the lexical items employed by members of KW, the consistency with Halliday's definition of anti-language becomes immediately apparent. There is a strong element of relexicalisation whereby all standard Reddit terminology has been modified to fit the needs and goals of the KW membership. Memetic lexical items of internet discourse in general have been replaced as well, as can be seen in the extensive use of calquing and other forms of language borrowing. There is even limited evidence of the over-lexicalisation common to anti-languages, which is exhibited by the various forms of the noun *upvotes* (e.g. *Aufwählsen*, *Hochwähl* and *Raufwahl*) and its related verb forms. This can also be seen in the differing forms of the calque

for *amirite*[28] of which three (*binichrechts*, *binichrichtig*, and *habichrecht*) are attested to in the keyword list.

The keyword list as in general indicates the extraordinary role which is played by a relatively tiny subset of lexical items. According to the list, just 2.17% (i.e. the top 1000 keywords) of all unique lexical items in the KW-Corpus-B data make up 17.85% of all tokens within the same. This demonstrates the importance of the specialised lexicon which is found in KW-discourse and not amongst the other subreddits which make up the GerRefSr. At the same time, this new vocabulary is largely built upon the same grammatical and syntactical foundations as the German found elsewhere on Reddit. Adjective are inflected according to the conventions of standard German, as are the conjugations of verbs, both novel and established. While certain novel nouns exhibit variation according to how their plural forms are marked, they all correspond to acceptable choices with German (e.g. *Maimais*, *Maimaier*, *Maimaien* (May-Mays[29])).

Also exhibited in the data, both from within the submission and lexical item data, is the extensive use of metaphor at all levels of KW-discourse, from the grammatical to the lexical, as can be seen in the following text taken from a submission (boldface is used to indicate the so-called title text of the submission):

[28] This is an alternative form of the phrase 'am I right' which mimics the speed at which it is commonly said, whereby the constituent lexemes are blended into a single word.

[29] *May-May* is a term used ironically as a substitute for the much more common *meme*. It is an orthographic attempt to match a well known mispronunciation of the term (Don, 2015).

Original:	Translation:
Volkszählung. Zur Teilnahme bitte Hochwählen. Wegwerfkonto weil es um die Wissenschaft geht und nicht ums Karma.	**Population Census. Upvotes please to take part.** Throwaway account because this is about science and not karma.

Grammatical metaphors[30] linked specifically to KW occur in the clause *Zur Teilnahme bitte Hochwählen* in which case there is a *logical metaphor* relating to the consequences of the (novel) verb *hochwählen*, which is in turn referred to metaphorically via use of the substantive. Meanwhile, *Wegwerfkonto* is the most obvious of several lexical metaphors, and is a semantic metaphor which equates the submitter's specific user account to something disposable which can be 'thrown away'.

While a comprehensive account of the use of metaphor within the KW-corpora is outside the scope of this thesis, looking at the keyword list reveals a great number of metaphoric lexical items which occur in nearly every possible part-of-speech, be it noun, verb, adjective, adverb, and so on. This strong integration of metaphor into KW-discourse is yet another consistency between it and the conceptualisation of anti-language.

I am of the opinion that the balance of evidence across all available data relating KW supports its classification using the framework laid out in Halliday (1976). This is classification is borne out by the oppositional nature of the KW-community as an anti-society within the larger German-language Reddit community, and Reddit more generally, as well as the metaphoric qualities of the language used in their discourse. This oppositional nature extends to the lexical items which are used specifically to avoid those which are not 'German enough' regardless of their accepted use in online contexts. The data with

[30]For a discussion relating to the concept of grammatical metaphor see Devrim (2015).

regards to widespread instances of relexicalisation and over-lexicalisation, as relatively limited as the over-lexicalisation may be in comparison to the examples provided by Halliday, also speaks to the language of KW as an anti-language. The end result is a new specialised subset of vocabulary which must be learnt in order to effectively integrate into KW 'society'; the discourse of KW is that built upon a foundation of an anti-language, which in turn forms the foundation of an anti-society.

6 Concluding Remarks and a Look forward

I have attempted to provide reasonable quantitative and qualitative evidence for a great number of things over the course of this thesis. My work was focused primarily on a single internet-based discourse-community, I endeavoured to do so using a linguistic concept first put forward by M.A.K. Halliday, anti-language, which to my knowledge has not yet been applied in this context. The theoretical foundation for this work was the theory of Systemic Functional Linguistics which was also initially developed by Halliday beginning in the early 1960s (Halliday & Matthiessen, 2013) and has since grown into a major theory guiding linguistic study and research.

In order to test my hypothesis that the key linguistic features of the KW-community represent those of anti-language as set forth in Halliday (1976), I have conducted a number of analyses on corpora and submission data. These analyses were undertaken in such a way that each provided a stepping stone for the next, each equally important for a reasonable determination to be made, be it positive or negative, of the presence of anti-language phenomena. All data analysed in this thesis was limited to the calender year 2015.

The initial step was a quantitative analysis of the vocabulary employed by KW members. Using the AntConc corpus analysis software developed by Laurence Anthony, in particular its keyword list generating algorithms, I sought to determine whether or not there were any significant lexical differences between KW and the greater German-language Reddit community as represented within the GerRefSr-corpora. This resulted in a list of 1648 key lexical items. Of those keywords, 1549 were found to have a statistical significance $\geq 50\%$ (EFFECT value ≥ 0.5000). These items represent only 3.37% of all lexical items within the KW-Corpus-B, but 25.82% of all tokens within said corpus. The average KEYNESS of these 1549 items was 316.91, while the

average EFFECT was 0.9141. The end result of this analysis is that there is a small but highly significant group of lexical items present in these KW-corpora data.

With the key lexical items identified, the next step was to determine their characteristics via a closer qualitative analysis. For this portion on the analysis, the top 1000 key lexical items were selected as determined by their EFFECT values. The average KEYNESS of these items was 428.66, with an average EFFECT values of 0.9824; these items represent 17.85% (66,763) of all tokens in the data. Upon analysis, eight categories of items were identified: ASCII Images; Bot-related Items; Existent Meanings; Image Macros; Loaned Items; Other-language Items; Subreddits; and Usernames. Of those eight categories, loaned items was the second largest in terms of unique items (247) after items with existent meanings consistent with Duden definitions (479). The loaned items were then selected for further analysis to determine the nature of their loaning. Calques represent the largest group of items, followed by loan-shifts, loan-blends, and imports. All but one loaned item was loaned from or through English, while the remaining item may possibly be linked to French.

After investigating the qualitative features of the KW-lexicon, I conducted a qualitative analysis of the contextual variables present in 55 user submissions to KW. This analysis was guided by Systemic Functional Linguistics, and the systems used were adapted from Wegener (2011) which in turn had been adapted from Butt (2004). Several layers of delicacy were also added by myself to highlight the subsystem features present in the submissions. The selection of submissions was based on variables relating to the submitting user and their karma scores. They were further organised into five distinct groups: Positively received submissions which are the only post on KW by the submitting user (11); poorly received submissions which are the only post on KW

by the submitting user (7); the top submissions of 2015 based on karma score (12); submissions with a karma score of zero (12); and a group of randomly selected submissions (12). The analysis determined that the most critical features correlating to positively received submissions were MEME within the SPHERE OF ACTION subsystem of the contextual variable FIELD, as well as CULTURAL CAPTIAL IN COMMON within the SOCIAL DISTANCE subsystem of the contextual variable TENOR. Negatively received submissions tended to exhibit the features SHITPOST or TROLL within the ACTION WITH SYMBOLS subsystem of FIELD, as was the feature NO CULTURAL CAPITAL IN COMMON. The contextual variable of MODE was largely unaffected given the constraints of Reddit as a medium, as were several subsystems within FIELD and TENOR. It was also determined that submissions featuring lexical items identified as part of the top 1000 tended to be associated with more positive receptions from the KW-community.

With the analyses of the KW-lexicon and discourse as it relates to user submissions completed, attention was turned to whether or not the balance of the evidence supports a conclusion that KW-discourse is indeed conducted in an anti-language. Both the overwhelming importance of lexical items key to KW, and their resistive nature corresponded to Halliday's conceptualisation of anti-language. This connection was further strengthened by the extensive evidence of relexicalisation relating to Reddit and web-terminology. There was also extensive use of metaphor present in both the lexical and the submission data, which Halliday (1976) identifies as a significant feature of anti-languages. Evidence of over-lexicalisation was present though to a far lesser extent than exhibited by anti-languages identified by Halliday (1976).

Ultimately, I have concluded the following: The lexicon of KW is unique compared to that of the larger German-language Reddit community; the use of

this lexicon is significant and has a tangible impact on the interaction between members of the KW-discourse community and visitors to the same; and finally, that given weight of all quantitative and qualitative evidence, the claim that the language of KW is the anti-language of an anti-society is justified.

The work in this thesis has been one which is a blend of the theoretical and the applied, the quantitative and the qualitative. The guiding thought behind it has been to provide a well-rounded analysis of a discourse community which I believe to be engaged in exciting linguistic activities. This carries with it a number of potential disadvantages and limitations in that it can quite easily lead to divided attention and limited depth of analysis. Admittedly, this could also be said of the present work; in looking at the lexicon of KW I restricted myself to only 1,000 out of nearly 46,000 items. Had I extended the scope of that particular analysis, it may well have provided an even more balance and nuanced account of that aspect of KW-discourse. The same could be said of my limited focus on submissions alone, leaving user comments unexamined. These I feel to be reasonable criticisms of the work. However, a truly comprehensive linguistic analysis of any given discourse community, regardless of size, is an enormous undertaking. Given the particular focus of my work, as well as more practical constraints relating to the format thereof, I must target my investigations more narrowly than may otherwise be desirable. In doing so I have attempted to address, at a minimum, the critical theoretical and conceptual considerations relevant to my research question. That being said, any errors of reason, logic, or application are mine alone, and should not be attributed to any presumed lack of robustness on behalf of the theories and concepts involved or the rigour with which they have been, and continue to be, developed.

Touching briefly upon a topic first mentioned in the introduction to this

work, what I was unable to truly elaborate on was the interplay between anonymity and language use. For one, KW is, of course, only semi-anonymous given that users assign themselves names which they can then develop into personae with unique histories and connections to the community. This system allows one (should they so chose) to separate their 'real' life from their interactions with and within KW. This may very well lead to behaviours and language use which would otherwise be avoided if not condemned by these same individuals in other situational contexts. Looking at the content of user submissions and comments on KW, one can certainly see how this may very well be the case. Unfortunately, however, due to the space and technical considerations made with regards to this work, I could only speculate as to the role which this has played within my data. I would all the same offer this gap in my analysis as an opportunity for real discovery about the roles and impacts of anonymity; as a semi-anonymous community, KW is at the cross-roads of these two states, the known and the unknown partner. Perhaps, given the correct analytical tools one can uncover how this potentially tenuous status as the person 'hidden' behind the user, influences the choice and manipulation of language and with them, the development of a discourse community.

I would also like to address the more technical limitations of my work. As the corpora in this thesis were developed from scratch within a very short time frame, certain aspects expected of professionally developed corpora are not found. The two which would quite likely have had the greatest positive impact on this work would have been part-of-speech tagging and the lemmatisation of the items found within the corpora. Given that all lexical comparisons between my corpora were conducted on equal terms, these elements would not change the overall outcomes of the analyses within this work. However, they would enable more effective and comprehensive research, and better provide a

stronger basis upon which to build an understanding of different communities involved.

My hope is that my thesis, despite any limitations in its scope or technique, has provided, if nothing else, an example that even seemingly trivial online phenomena and discourse is fertile ground for linguistic investigation. Reddit and the RPDA represent a massive treasure chest of linguistic data, sourced from around the world, not only in English and German, but a multitude of languages. For the corpus linguists among us, the RPDA represents one of the largest corpora available, English or otherwise, and there are numerous opportunities to use it to build or supplement more specialised corpora; there are thousands of discourse communities to be analysed. For those whose academic interests are geared towards computer mediated communication, or communication more broadly, you are presented with a fantastic wealth of observable phenomena, while those looking to engage in multi-modal analyses could hardly ask for a more diverse collection of modalities in an online context.

If this work has done nothing else than pique curiosity around the ideas of the CJ and SP, then all the better, as they too represent phenomena worthy of a closer look. KW too, is ripe for further study; what role does reputation play in the reception of submissions? What has been the impact of KW discourse on that of the other German subreddits? Can their influence be seen elsewhere on Reddit or perhaps even other areas of the internet? These are just some of the questions which are a left unanswered here, but await researchers should they wish to seek them out.

References

Al-Azmeh, A. (1991). Islamist Revivalism and Western Ideologies. *History Workshop*(32), 44-53.

Alexa. (2016). *reddit.com Site Overview.* Retrieved 2016-2-19, from `http://www.alexa.com/siteinfo/reddit.com`

Anonymous. (2011). *fappieren.* Retrieved 2016-6-15, from `http://www.sprachnudel.de/woerterbuch/fappieren`

Anthony, L. (2016). *AntConc.* Retrieved from `http://www.laurenceanthony.net`

B, A. (2012). *Karma Whore.* Retrieved 2016-3-23, from `http://knowyourmeme.com/memes/karma-whore`

Barthel, M., Stocking, G., Holcomb, J., & Mitchell, A. (2016). *Nearly Eight-in-Ten Reddit Users Get News on the Site.* Retrieved 2016-4-28, from `http://www.journalism.org/files/2016/02/PJ_2016.02.25_Reddit_FINAL.pdf`

Baumgartner, J. (2015). *Complete public reddit comments corpus.* Retrieved 2016-07-17, from `https://archive.org/details/2015_reddit_comments_corpus`

Baumgartner, J. (2016). Personal Communication.

Beier, A. L. (2005). Identity, Language, and Resistance in the Making of the Victorian Criminal Class: Mayhew's Convict Revisited. *Journal of British Studies*, *44*(3), 499-515.

Benson, J. D., & Greaves, W. S. (Eds.). (1988). *Systemic functional perspectives on discourse.* Norwood N.J.: Albex.

Berger, P. L. (1966). *The social construction of reality : a treatise in the sociology of knowledge* (1st ed., Vol. 78) (No. 3). Garden City, N.Y.: Doubleday.

Boddy, J. P. (1989). *Wombs and alien spirits women, men, and the Zār cult in northern Sudan*. Madison, Wis.: University of Wisconsin Press.

Bolton, K., & Hutton, C. (1995). Bad and Banned Language: Triad Secret Societies, the Censorship of the Cantonese Vernacular, and Colonial Language Policy in Hong Kong. *Language in Society*, *24*(2), 159-186.

Butt, D. G. (2004). *Parameters of context.* (Mimeograph)

Certeau, M., & Brammer, M. (1992). Mysticism. *Diacritics*, *22*(2), 11-25.

Champagne, L. (1977). The Beach beneath the Paving Stones: May 1968 and French Theater. *SubStance*, *6-7*(18), 59-71.

Chesney, K. (1970). *The anti-society: an account of the Victorian underworld*. Gambit.

Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.

De Lauretis, T. (1988). Sexual Indifference and Lesbian Representation. *Theatre Journal*, *40*(2), 155-177.

Devrim, D. (2015). Grammatical metaphor: What do we mean? what exactly are we researching? *Functional Linguistics*, *2*(1), 1-15.

Domeris, R., William. (1999). When metaphor becomes myth: A sociolinguistic reading of Jeremiah. In A. R. P. Diamond, K. M. O'Connor, & L. Stulman (Eds.), (p. 244-262). Sheffield, England: Sheffield Academic Press.

Don. (2013). *Circle Jerk.* Retrieved 2016-3-23, from http://knowyourmeme.com/memes/circle-jerk

Don. (2015). *Maymay.* Retrieved 2016-6-13, from http://knowyourmeme.com/memes/maymay

Duggan, M., & Smith, A. (2013). *6% of Online Adults are Reddit Users.* Retrieved 2016-4-28, from http://www.pewinternet.org/files/old-media//Files/Reports/2013/PIP_reddit_usage_2013.pdf

Fawcett, R. P. (2009). The Many Types of 'Theme' in English: their Syntax, Semantics and Discourse Functions.

Fawcett, R. P. (2013). Choice and choosing in Systemic Functional Grammar. In L. Fontaine, T. Bartlett, & G. O'Grady (Eds.), (1st ed., p. 115-134). Cambridge, England: Cambridge University Press.

Fontaine, L. (2013). Introduction: choice in contemporary systemic functional theory. In L. Fontaine, T. Bartlett, & G. O'Grady (Eds.), (1st ed., p. 1-12). Cambridge, England: Cambridge University Press.

Fontaine, L., Bartlett, T., & O'Grady, G. (2013). *Systemic functional linguistics: Exploring choice.* Cambridge, England: Cambridge University Press.

freshoutofthebox. (2013). *r/SCHLAND Frontpage.* Retrieved 2016-7-05, from https://de.reddit.com/r/schland

Grimm, J., & Grimm, W. (1862). *Deutsches Wörterbuch* (Vol. 1). Leipzig: Hirzel.

Groenewald, J. (2011). Show, tell and re- enact: The reason why the earliest followers of Jesus found the Eucharist meaningful. *HTS Teologiese Studies/Theological Studies*, *67*(1).

Halliday, M. A. K. (1967). Notes on Transitivity and Theme in English: Part 1. *Journal of Linguistics*, *3*(1), 37-81.

Halliday, M. A. K. (1976). Anti-languages. *American Anthropologist*, *78*(3), 570-584.

Halliday, M. A. K. (1978). *Language as social semiotic : the social interpretation of language and meaning*. Baltimore: University Park Press.

Halliday, M. A. K. (2009). Language as Social Semiotic: Towards a General Sociolinguistic Theory. In (Vol. 10, p. 169-201). London; New York, NY: Continuum.

Halliday, M. A. K., & Matthiessen, C. M. (2013). *Halliday's introduction to functional grammar* (4th ed.). New York, N.Y.: Routledge.

Halliday, M. A. K., & Webster, J. J. (2009). *Continuum companion to systemic functional linguistics*. London ; New York, N.Y.: Continuum.

Harris, R. (1997). Possession on the Borders: The "Mal de Morzine" in Nineteenth-Century France. *The Journal of Modern History*, *69*(3), 451-478.

Hasan, R. (1977). Text in the Systemic-Functional Model. In W. U. Dressler (Ed.), (p. 228-246). Berlin: Walter de Gruyter.

Hasan, R. (1996). *Ways of saying: ways of meaning: selected papers of Ruqaiya Hasan* (C. Cloran, Ed.). London: Cassell.

Hasan, R. (1999). Speaking with Reference to Context. In M. Gahdessy (Ed.), (p. 219-328). Amsterdam: John Benjamins.

Hasan, R. (2013). Choice, system, realisation: describing language as meaning potential. In (1st ed., p. 269-299). Cambridge, England: Cambridge University Press.

Hasan, R. (2014). Towards a paradigmatic description of context: systems, metafunctions, and semantics. *Functional Linguistics*, *1*(1), 1-54.

Heath, D. (1992). Fashion, Anti-Fashion, and Heteroglossia in Urban Senegal. *American Ethnologist*, *19*(1), 19-33.

Hockett, C. F. (1958). *A course in modern linguistics.* New York, Macmillan 1958.

Hutcheon, L. (1992). The complex functions of irony. *Revista Canadiense de Estudios Hispnicos*, *16*(2), 219-234.

kharanos, & noob, l. (2013). *Warlizard Gaming Forum.* Retrieved 2016-6-10, from `http://knowyourmeme.com/memes/warlizard-gaming-forum`

Killerdildo. (2010). *fappen.* Retrieved 2016-6-15, from `http://www.mundmische.de/bedeutung/26218-fappen`

Kohn, L. (2008). Antilanguage and a Gentlemans Goloss: Style, Register, and Entitlement to Irony in A Clockwork Orange. *eSharp*, *11*, 1-27.

Leech, G., & Fallon, R. (1992). Computer corpora-what do they tell us about culture. *ICAME journal*, *16*, 29-51.

Lehman, W. (1968). Crime, the Public, and the Crime Commission: a Critical Review of "The Challenge of Crime in a Free Society," The Report of the President's Commission on Law Enforcement and the Administration of Justice. *Michigan law review*, *66*, 1487-1540.

Martin, J. (1991). Intrinsic functionality: Implications for contextual theory. *Social Semiotics*, *1*(1), 99-162.

Martin, J. (2009). Realisation, Instantiation and Individuation: Some Thoughts on Identity in Youth Justice Conferencing. *Revista de Documentacao de Estudos em Linguistica Teorica e Aplicada (D.E.L.T.A.)*, *25*, 549-583.

Martin, J. R. (1995). Text and clause: Fractal resonance. *Text - Interdisciplinary Journal for the Study of Discourse*, *15*(1), 5-42.

Matthiessen, C. M. (1988). Representational issues in systemic functional grammar. In J. D. Benson & W. S. Greaves (Eds.), (p. 221-292). Norwood N.J.: Albex.

Matthiessen, C. M. (2009). Ideas and new directions. In (1st ed., p. 12-58). London; New York, NY: Continuum.

Matthiessen, C. M. (2012). Systemic Functional Linguistics as appliable linguistics: social accountability and critical approaches. *DELTA: Documentao de Estudos em Lingstica Terica e Aplicada*, *28*, 435-471.

Mcintosh, J. (2004). Reluctant Muslims: Embodied hegemony and moral resistance in a giriama spirit possession complex. *Journal of the Royal Anthropological Institute*, *10*(3), 91-112.

McNerney, W. J. (1969). Changing the Health Care System. *The American Journal of Nursing*, *69*(11), 2428-2435.

Mercer, A., & Zed, T. (2014). *Shitposting.* Retrieved 2016-3-21, from `http://knowyourmeme.com/memes/shitposting`

Merrim, S. (1980). A Secret Idiom: The Grammar and Role of Language in "Tres Tristes Tigres". *Latin American Literary Review*, *8*(16), 96-117.

Morris, & Brad. (2014). *Press F to Pay Respects.* Retrieved 2016-6-15, from http://knowyourmeme.com/memes/press-f-to-pay-respects

Nico Jacobellis v. Ohio, 378 U.S. 184 (1964)

NovaXP, & Asdfghjkl. (2012). *Lenny Face.* Retrieved 2016-6-28, from http://knowyourmeme.com/memes/lenny-face

Onysko, A. (2007). *Anglicisms in German: Borrowing, Lexical Productivity, and Written Codeswitching.* Walter de Gruyter.

Oxman, N. (2010). Per formative: Toward a Post-Formal Paradigm in Architecture. *Perspecta*, *43*, 19-177.

Podgrecki, A. (1973). *"Second Life" and its Implications* (Vol. 78) (No. 3). Wiley Online Library. (Mimeograph)

Posteguillo-Gómez, S. (2002). Netlinguistics and English for Internet Purposes. *Ibrica: Revista de la Asociacin Europea de Lenguas para Fines Especficos (AELFE)*(4), 21-38.

Pre.ethics.gc.ca. (2014). *2. scope and approach: The interagency advisory panel on research ethics (pre).* Retrieved 2016-2-19, from http://www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/chapter2-chapitre2/

Reddit. (2015a). *Frequently Asked Questions.* Retrieved 2016-3-21, from https://www.reddit.com/wiki/faq

Reddit. (2015b). *reddit.com: about reddit.* Retrieved 2016-2-19, from https://www.reddit.com/about

Reddit. (2016a). *reddit.com: api documentation.* Retrieved 2016-2-19, from https://www.reddit.com/dev/api

Reddit. (2016b). *reddit.com: privacy policy.* Retrieved 2016-2-19, from https://www.reddit.com/help/privacypolicy

Reddit. (2016c). *Reddit Gold.* Retrieved 2016-3-21, from https://www.reddit.com/gold/about/

Redditmetrics. (2016a). *New subreddits by date.* Retrieved 2016-07-17, from http://redditmetrics.com/history

Redditmetrics. (2016b). */r/kreiswichs metrics (KREISWICHS - Ente Ente Ente Ente Ente Ente Ente Ente Ente).* Retrieved 2016-2-19, from http://redditmetrics.com/r/kreiswichs

Redditmetrics. (2016c). */r/SCHLAND ('SCHLAND OLEEEEE).* Retrieved 2016-2-19, from http://redditmetrics.com/r/SCHLAND

Rediker, M. (1981). "Under the Banner of King Death": The Social World of Anglo-American Pirates, 1716 to 1726. *The William and Mary Quarterly*, *38*(2), 203-227.

Reeves, E. B. (1995). power, resistance, and the cult of muslim saints in a northern Egyptian town. *American Ethnologist*, *22*(2), 306-323.

Research. (2014). *Does my data collection activity require ethics review?* Retrieved 2016-2-19, from https://uwaterloo.ca/research/office-research-ethics/research-human-participants/pre-submission-and-training/human-research-guidelines-and-policies-alphabetical-list/does-my-data-collection-activity-require-ethics-review

Ridge, E., & Kudenko, D. (2010). Tuning an Algorithm Using Design of Experiments. In T. Bartz-Beielstein, M. Chiarandini, L. Paquete, & M. Preuss (Eds.), *Experimental methods for the analysis of optimization algorithms.*

Rosenfeld, B., & Penrod, S. (2011). *Research Methods in Forensic Psychology.* Wiley.

Salihefendic, A. (2015). *How Reddit ranking algorithms work.* Retrieved 2016-3-21, from `https://medium.com/hacking-and-gonzo/how-reddit-ranking-algorithms-work-ef111e33d0d9#.4vl4nkqo6`

Schniedewind, W. M. (1999). Qumran Hebrew as an antilanguage. *Journal of Biblical Literature*, 235-252.

Sprott, S. E. (1969). The Damned Crew. *PMLA*, *84*(3), 492-500.

Staller, N. (1994). Babel: hermetic languages, universal languages, and anti-languages in fin de siècle parisian culture. *The Art Bulletin*, *76*(2), 331-354.

steve. (2006). *it's about time we added something new.* Retrieved 2016-07-17, from `http://www.redditblog.com/2006/02/its-about-time-we-added-something-new.html`

Stewart, S. (1980). The Pickpocket: A Study in Tradition and Allusion. *MLN*, *95*(5), 1127-1154.

Stimpson, C. R. (1985). The Somagrams of Gertrude Stein. *Poetics Today*, *6*(1), 67-80.

Stuck_In_The_Matrix. (2015). *I have every publicly available Reddit comment for research. ~1.7 billion comments 250 GB compressed. any interest in this? /r/datasets.* Retrieved 2016-

2-19, from https://www.reddit.com/r/datasets/comments/3bxlg7/i_have_every_publicly_available_reddit_comment

Tatham, S. (2016). *PuTTY: a free SSH and Telnet client.* Retrieved 2016-6-27, from http://www.chiark.greenend.org.uk/~sgtatham/putty/

Tomberry. (2010). *Fap.* Retrieved 2016-3-21, from http://knowyourmeme.com/memes/fap

Tseng, M.-y. (1997). Symbolic discourse: mystical writing as anti-language. *Language and Literature*, *6*, 181-196.

Villoria, M. (2011). Colombia's drug trafficking subculture: its literary representation in La Virgen de los sicarios and Rosario Tijeras.(critical essay). *Caribbean Quarterly*, *57*(2), 75-91.

Webster, J. J. (2009). An Introduction to Continuum Companion to Systemic Functional Linguistics. In (1st ed., p. 1-11). London ; New York, N.Y.: Continuum.

Wegener, R. (2011). *Parameters of context: from theory to model and application* (Doctoral dissertation, Macquarie University, Sydney). Retrieved from http://www.isfla.org/Systemics/Print/Theses/Wegener_PhD_Print_Version.pdf

Wikipedia. (2016a). *Json.* Retrieved 2016-07-06, from https://en.wikipedia.org/wiki/JSON

Wikipedia. (2016b). *Reddit.* Retrieved 2016-2-20, from https://en.wikipedia.org/wiki/Reddit

Subsystem Network Diagrams & Subreddit Frontpages

The Subsystems within the Contextual Variable of Field

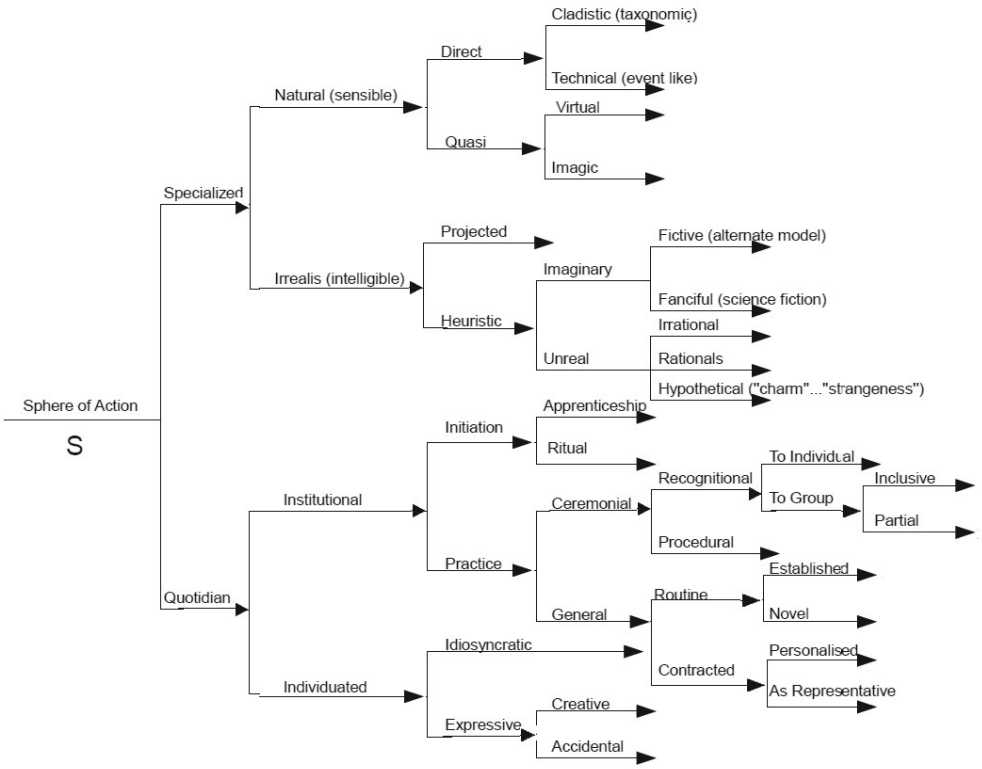


Figure 1: The SPHERE OF ACTION subsystem. Adapted from Wegener (2011).

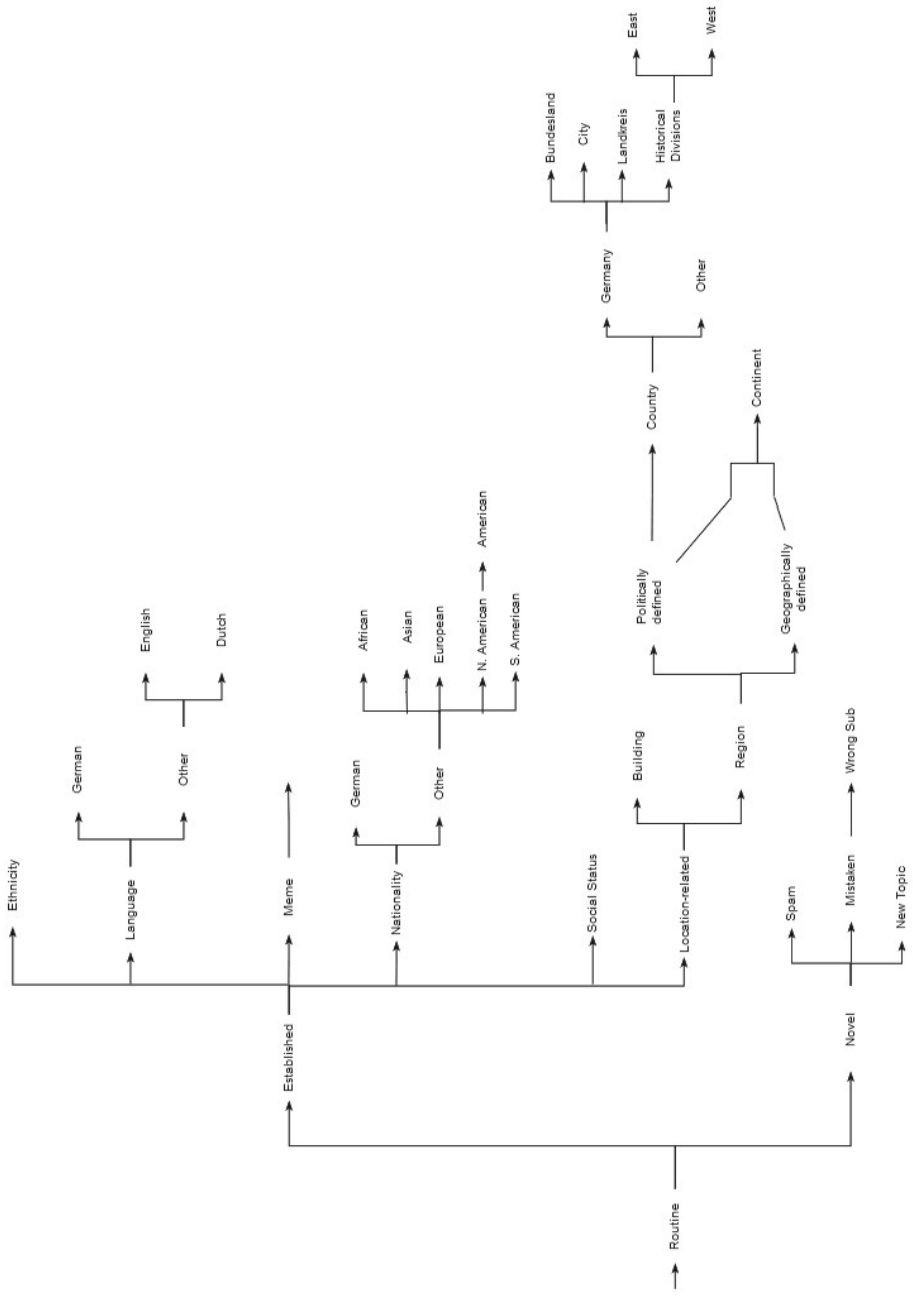


Figure 2: Feature network extending in delicacy from the feature ROUTINE in the SPHERE OF ACTION subsystem.

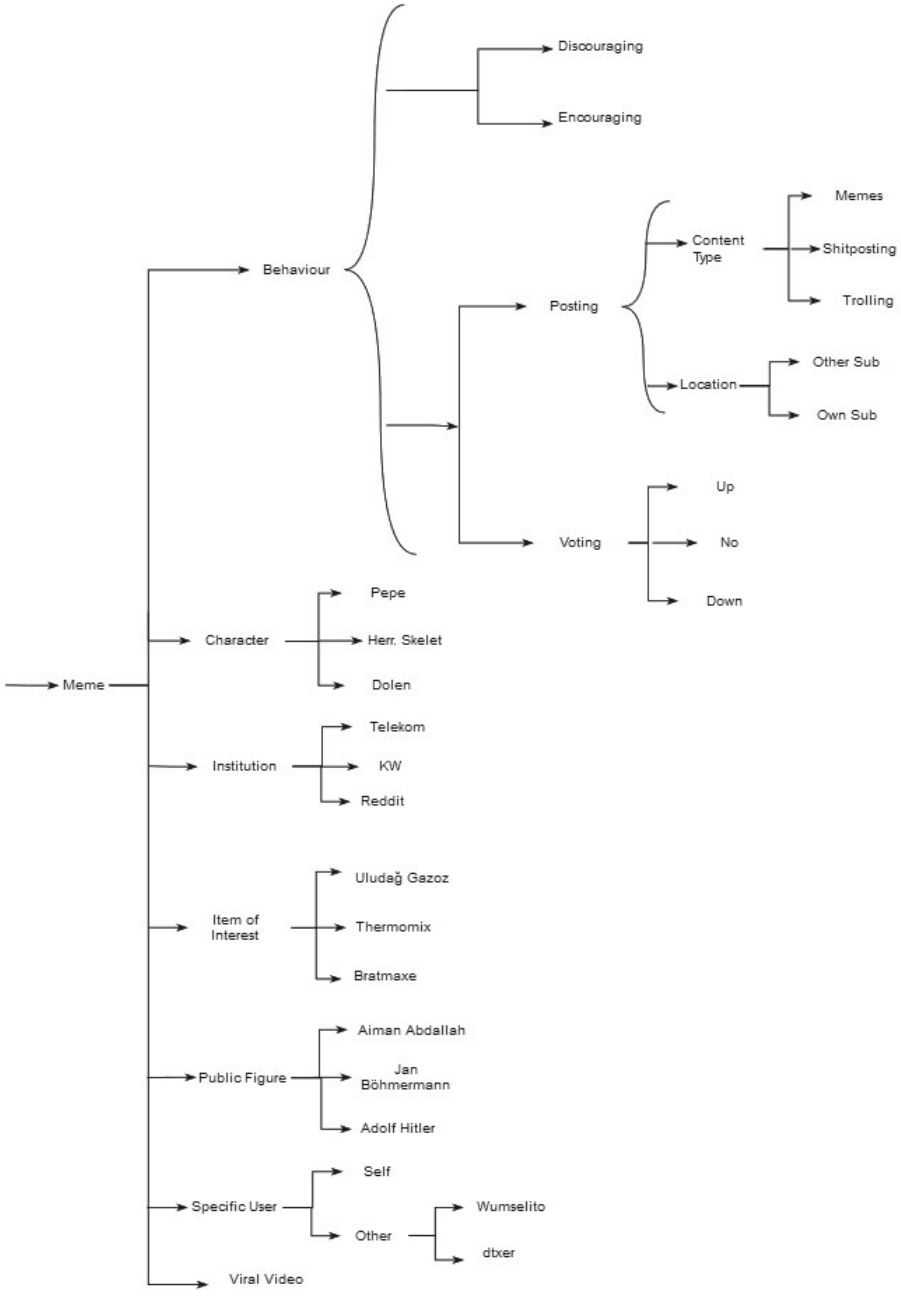


Figure 3: Feature network extending in delicacy from MEME in the SPHERE OF ACTION subsystem.

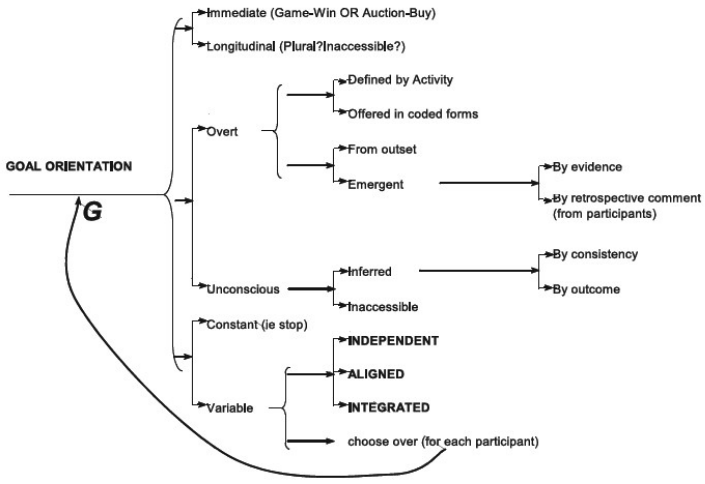


Figure 4: The GOAL ORIENTATION subsystem. Adapted from Wegener (2011).

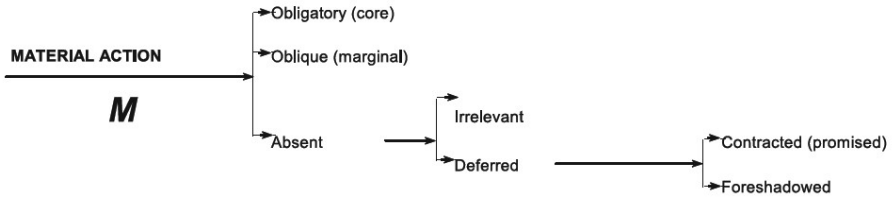


Figure 5: The MATERIAL ACTION subsystem. Adapted from Wegener (2011).

Figure 6: The ACTION WITH SYMBOLS subsystem. Adapted from Wegener (2011).

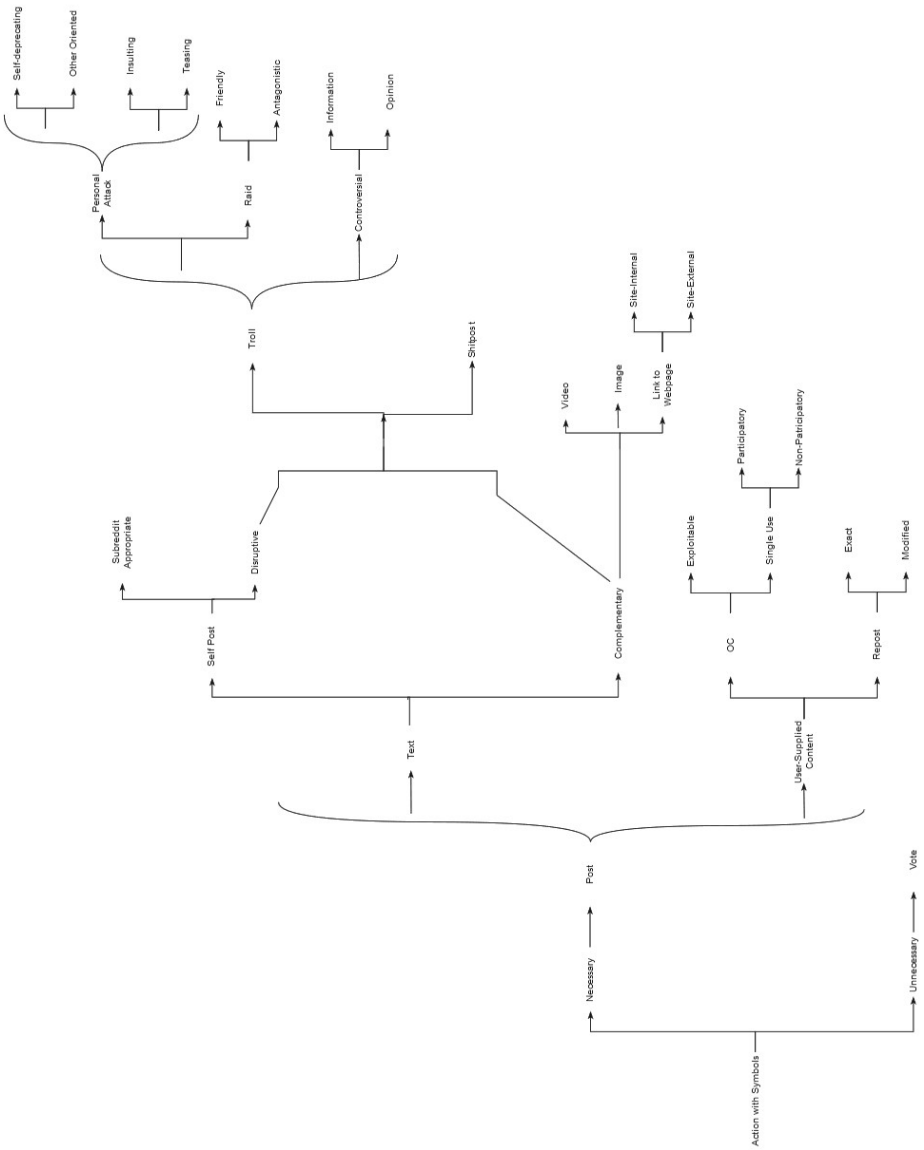


Figure 7: Extensions of the ACTION WITH SYMBOLS subsystem.

The Subsystems within the Contextual Variable of Tenor

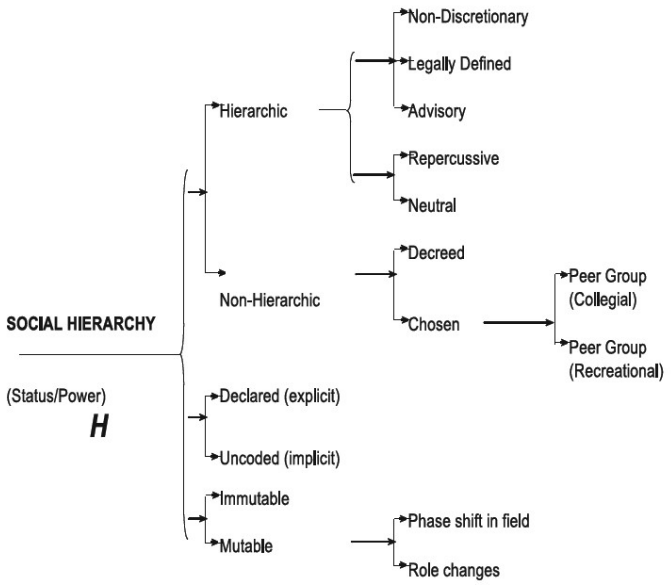


Figure 8: The feature network of the SOCIAL HIERARCHY subsystem. Adapted from Wegener (2011).

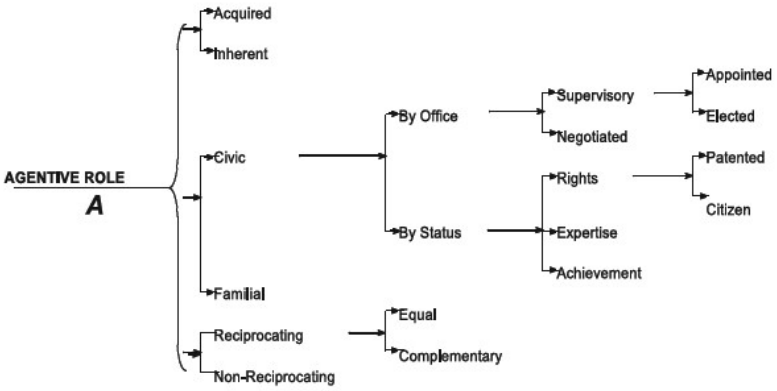


Figure 9: The feature network of the AGENTIVE ROLE subsystem. Adapted from Wegener (2011).

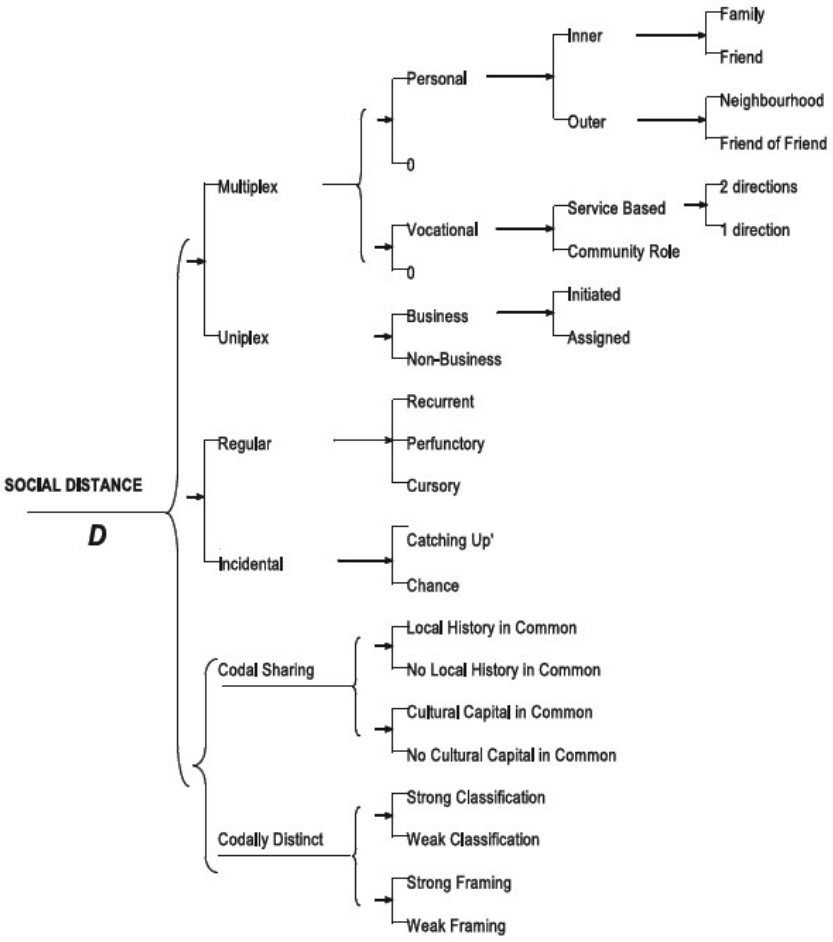


Figure 10: The feature network of the SOCIAL DISTANCE subsystem. Adapted from Wegener (2011).

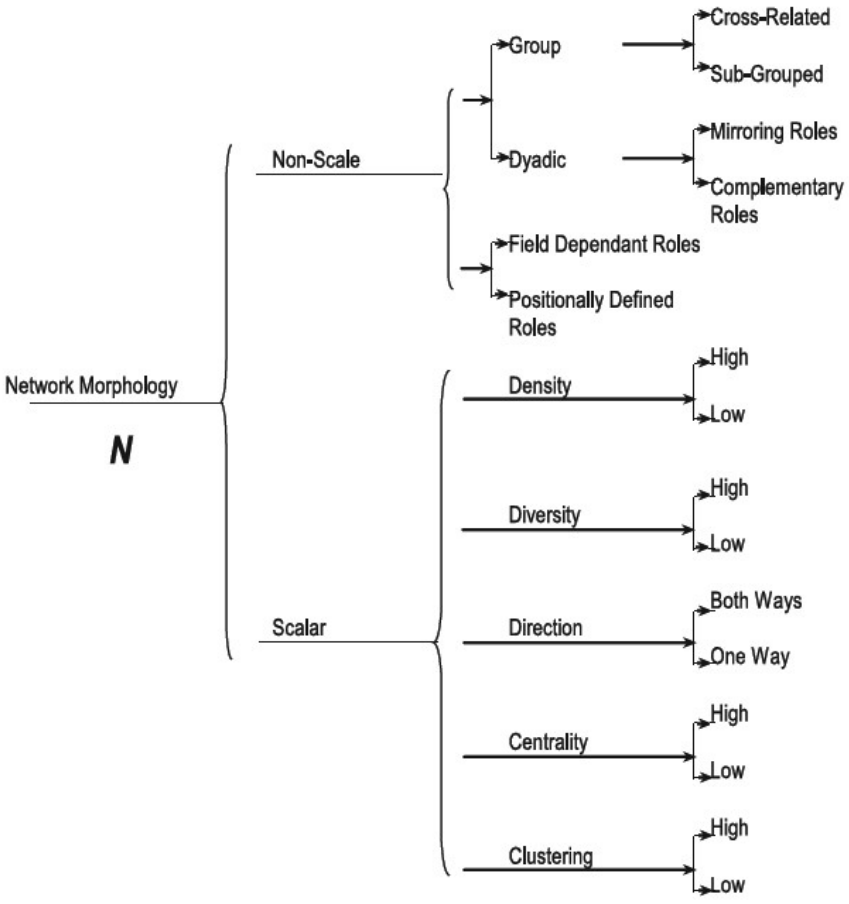


Figure 11: The feature network of the NETWORK MORPHOLOGY subsystem. Adapted from Wegener (2011).

The Various Frontpages of the Subreddits discussed in this Thesis

Figure 12: The frontpage of r/de.

Figure 13: The frontpage of r/SCHLAND.

Figure 14: The frontpage of r/self_de.

Figure 15: The frontpage of r/kreiswichs.