From Frequency to Sequence: How Quantitative Methods Can Inform Qualitative Analysis of Digital Media Discourse

Dang-Anh, Mark; Rüdiger, Jan Oliver

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

From Frequency to Sequence: How Quantitative Methods Can Inform Qualitative Analysis of Digital Media Discourse

(Journal Article + Infographic)

Mark Dang-Anh |
Jan Oliver Rüdiger

This paper aims at showing how quantitative corpus linguistic analysis can inform qualitative analysis of digital media discourse with respect to the mediality of language in use. Using the example of protest discourse in Twitter, in the field of anti-Islamic 'Pegida' demonstrations, a three-step method of collecting, reducing and interpreting *salient* data is proposed. Each step is aligned with operative medial features of the microblog: hashtags, retweets and @-interactions. The exemplary analysis reveals the importance of discussions of attendance numbers in protest discourse and the asymmetry between administrative (i.e. the police) and non-administrative discourse agents. Furthermore, it exemplifies how frequency analysis and sequence analysis can be combined for research in media linguistics.

This contribution also comprises an *infographic* which can be retrieved from www.10plus1journal.com.

10PLUS1
LIVING
LINGUISTICS

1. Introduction

In todays' digitally mediatized world, micro-blogging is a common practice in political communication (Thimm et al. 2014). Twitter, as the most popular microblogging platform in the western hemisphere[1], is a medium that is used widely by political actors, be it on an institutional or individual level. The course and discourse of street protests, especially, is constituted by microblog communication (Gerbaudo 2012). Although there are numeral studies on the role of digital media in street protests, be it for example in the context of the Occupy protests (Penney & Dadas 2014), the Tahrir Square protests (Wilson & Dunn 2011; Tufekci & Wilson 2012), or anti-fascist protests (Dang-Anh & Eble 2013; Neumayer & Valtysson 2013), few take a perspective on the linguistic construction of meaning for and within these streets protests. Furthermore, linguistic explorations dealing with language and protest (cf. Martín Rojo 2014a) focus on on-street/ square semiotics, e.g. on banners (Martín Rojo 2014b), and offline discourse of street assemblies (Steinberg 2014). As such, they don't consider the polyphonic occurrences of language in social media platforms such as Facebook, Twitter, Instagram, and so forth.

This paper aims at bridging the delineated gap between linguistic and medial foci with an emphasis on the methodological and methodical question of how corpus linguistics can inform qualitative inquiries in media linguistics. In other words: Media linguistics can incorporate quantitative corpus linguistics and qualitative hermeneutic analysis with respect to the mediality of language use, as every processed occurrence of language has its own mediality, i.e. features of media including the opportunities and constraints they impose on communicative practices (and vice versa). The case of Twitter usage in street protest thus serves as an exemplification of how quantitative and qualitative methods might be sensibly combined.[2]

Given the fact that large numbers of people contribute to digital media discourses on political events, such as protests, and thus create large numbers of texts, research methods must be adapted to these phenomena for the purpose of linguistic analysis.

[1] Twitter is banned in China. The most popular micro-blogging platform in China is Sina Weibo.

[2] For illustration, this article is complemented with an **infographic** which is available at *10plus1journal.com*.

However, the sheer vastness of communicative occurrences of language in digital media does not prevent linguistic analysis from the hermeneutic (re-)construction of meaning through in-depth analysis. Thus, it is our goal to show how corpus linguistics as a quantitative means can heuristically inform qualitative analysis of digital media discourse. To be more precise, we will show how the analysis of *frequencies* structures and filters large datasets in order to practically perform qualitative analysis of communicative *sequences*. By identifying *salient* utterances and speakers through quantitative measures the three-step method of data collection, reduction, and interpretation points to a communicative sequence, the *meaning* of which is qualitatively analysed with respect to its context. However, our approach is not strictly linear but circular, as we reassess the quantitative analytical steps by corpus-hermeneutic reflections.[3]

The following section outlines the relation between quantitative and qualitative methods with special regard to the notion of *salience* in conjunction with those of frequency and ascriptions of relevance. The next section introduces Twitter, the operativity of hashtagging, retweeting and @-mentioning and their alignment with the proposed three-step method of analysis. Following this, the method is exemplified using the case of 'Pegida' protests and a corresponding Twitter corpus. We conclude both, the methodological reflection and the exemplary analysis, in the last section.

2. Corpus Linguistics as Heuristics for Qualitative Analysis

When it comes to researching large data sets, it is increasingly regarded as sensible and fruitful to extend qualitative methods by quantitative means (cf. O'Halloran 2010; O'Keeffe 2012; Bubenhofer 2013; Baker & Levon 2015). McEnery & Hardie, thus taking a stance on corpus linguistics as a supportive extension of qualitative research methods: "any field that is based, primarily or in part, on the study of text can benefit from corpus methods in any research context where the body of text that is of interest expands beyond the point where hand-and-eye methods of analysis can fully encompass its contents." (2012: 231) In their work on written discourse, Cameron & Panović emphasize the advantage of the quantity of 'evidencing' data: "by using computer software, analysts can deal with much larger quantities of data, and so put forward more convincing evidence in support of their claims." (2014: 81) Both views emphasize the utility of corpus linguistic methods for the analysis of large data sets. However, data and analyses based on data, be it from small or large corpora, must always be interpreted and contextualized in order to conduct an adequate reconstructive analysis of the construction of meaning within social contexts. From such a perspective, *evidentiality*[4] or *cogency* is not an inherent feature of a datum but *evidencing* is a reflexive scientific process that data analysts negotiate intersubjectively. As a consequence, quantitative and qualitative methods have to be balanced precisely. While qualitative methods still pave the royal road to reconstructive analysis of language-in-use, quantitative methods might support analyses heuristically. By accessing large data sets heuristically with the use of corpus linguistics, subtle communicative patterns might emerge, intuitively expected commu-

[3] This is exemplified in the first step of the exemplary analysis (cf. p.64).

[4] *Evidentiality* is *not* used here in its specific linguistic terminological sense that points to grammatical features (cf. Chafe & Nichols 1986).

nicative patterns might be confirmed and thus the foci of qualitative analyses can be directed to distinct pieces of data.[5]

3. Salience and Frequency

We use *salience* as a key concept to identify distinct communicative sequences from large data sets. Klein (2014) conceptualizes a trans-situational understanding of "salient sentences" by which he refers to epoch-defining utterances[6] such as the first sentence of the Basic Law for the Federal Republic of Germany ("Human dignity shall be inviolable."; Basic Law for the Federal Republic of Germany 2012: 15) or sentences that provoke political discussion, such as Germany's former president Wulff's utterance in a speech: "Islam belongs to Germany."[7] Salience, in this sense, points to highly visible acts of political communication that characterize political systems or coin political discourse over a trans-situational time span. However, Klein acknowledges that salient sentences necessarily belong to a collective actual knowledge and possibly anchor within the collective memory of politically and historically interested parts of society (Klein 2014: 122). From Klein's (2014: 123) point of view, three features are required for sentences to become salient:

1) "a considerable speaker (person, groupings)"
2) "a politically relevant topic (under the conditions of a democratic system, this predominatly means a controversial topic)"
3) "a special situation that is evoked by public attention, aggravation, or arousing a latent controversy".

As the first two – rather vaguely attributed – features of considerability and political relevance indicate, Klein clearly has in mind larger political issues with their own historicity. Contrary to such discourses mainly driven by mass media, in digital media discourse, the first two requirements are not "indispensable" (Klein 2014: 123), as speakers and topics might emerge as well, i.e. postings from non-prominent speakers about non-current topics might trigger public debates. However, even for singular events – such as those discussed here that occurred in a series of iterative protest phenomena under the label of *PEGIDA* – the notion of salience is helpful for our approach to identify relevant speakers and utterances within a corpus of thousands of tokens. In a broader sense that emphasizes the aspect of perceptivity from the perspective of mass communication research, salience is understood as the act of "making a piece of information

---

[5] However, this merely describes *one way* amongst others to *cherry-pick* the focus of qualitative text analyses (cf. Baker & Levon 2015).

[6] As *sentence* is a categorization mainly for written language, we prefer the term *utterance* from a pragmatical and media linguistic point of view in order to overcome the spoken-written-dichotomy and to prevent from a "written language bias" (Linell 1982).

[7] This reproduction is slightly shortened. Wulff originally uttered: "Das Christentum gehört zweifelsfrei zu Deutschland. Das Judentum gehört zweifelsfrei zu Deutschland. Das ist unsere christlich-jüdische Geschichte. Aber der Islam gehört inzwischen auch zu Deutschland." In English (our translation): "Christianity belongs without any doubt to Germany. Judaism belongs without any doubt to Germany. But Islam meanwhile also belongs to Germany." The translated manuscript from his speech on Reunification Day in Bremen, 3rd October 2010 is available on the German parliament's website (cf. Wulff 2010). Interestingly, the official translation adds the notion of 'identity' to Wulff's utterance: "Christianity is without a doubt part of German identity. Judaism is without a doubt part of German identity. Such is our Judaeo-Christian heritage. But Islam has now also become part of German identity." These nuanced translational differences create differences in meaning that cannot be further discussed here. However, the deliberate defusing of the original utterance by official authorities underscores the salience of Wulff's utterance in Klein's sense.

more noticeable, meaningful, or memorable to audiences. An increase in salience enhances the probability that receivers will perceive the information […]" (Entman 1993: 53). Despite arguing with the most problematic concepts from (mass) communication studies like *audience* and *receiver*, one can transfer the concept of salience to Twitter, where the relation between the original poster, the retweeter and the recipient (of the retweet) is cascading as the latter can herself become a retweeter (and so forth). As such, making a posting relevant by retweeting it is a productive practice whereas perceiving relevance and thus the feature of a posting as salient is a merely receptive one.

Consequentially, Klein identifies *resonance* as the key factor of responsive ascriptions of salience to utterances:

> Whether a sentence is apprehended, spread and finally becomes commonly known depends on the resonance amongst the recipients, especially leading political media […]. Here, quantity (medial distribution) and intensity (degree of accentuation, citing or referring) play a central role (Klein 2014: 123).[8]

The understanding of utterances that are frequently made relevant (and thus become salient) in the course of their interpretative reception yields methodological implications for the field of salient sentences: "For its systematic processing the conjunction of corpus linguistics and linguistic hermeneutics might be the appropriate methodical approach." (Klein 2014: 125).[9] Consequently, computational procedures of corpus linguistics and intersubjectively scrutinized qualitative-hermeneutical methods should not be viewed as contradictory but as complementary (Felder 2012: 125).

In social media platforms, contrary to the production of resonance by gatekeepers in mass media discourse, it is the recipients who make other agent's postings more relevant by retweeting them.[10] Whereas agents ascribe relevance to postings by singular acts of retweeting, the accumulation of retweets makes this relevance more perceivable – as indicated by the retweet count under each posting – and as such salient. Salience is thus established by reflexive "quantifiable valuation practices" (Paßmann 2015: 141) that social media platforms make possible through their medial features. As a consequence, the visibility of practices in social media platforms makes them accessible for research: "Platform activities [such as retweets] directly connect practices with data that are generated in the process of retweeting. Thus, user activities become summable." (Paßmann & Gerlitz 2014: 2)[11] Taking into account the reflexivity of social media data (Paßmann 2014) and the agency of social media users that finds expression in the communicative practices of social media users, the categories of *considerable speakers* and *politically relevant topics* in social media discourse are not historical in Klein's sense

[8] In this paper, all German citations are translated into English. Original: „Ob ein Satz aufgegriffen, weitergetragen und schließlich allgemein bekannt wird, hängt aber von der Resonanz bei den Rezipienten ab, insbesondere bei den politischen Leitmedien (die großen TV-Sender, überregionale Zeitungen und politische Magazine in Print- und Online-Format). Dabei spielen Quantität (mediale Distribution) und Intensität (Grad der Hervorhebung, Zitieren oder Referieren) ein zentrale Rolle. (Klein 2014: 123)

[9] Original: "Für seine systematische Bearbeitung dürfte die Verknüpfung von Korpuslinguistik und linguistischer Hermeneutik der angemessene methodische Ansatz sein.“ (Klein 2014: 125)

[10] It is important to note that recent studies in media research point to the fact that the distribution of postings in social media platforms is selective (Dang-Anh et al. 2013b) and thus platforms like Twitter Inc. (Halavais 2014) have their own agenda, doing *platform politics* (Gillespie 2010).

[11] Original: "Indem Plattformaktivitäten eine direkte Verbindung zwischen Praktiken und den dabei erzeugten Daten herstellen, werden Daten von Nutzeraktivitäten aggregierbar.“ (Paßmann and Gerlitz 2014: 2).

but are ascribed perceivable relevance and, as such, salience is established through user practices.

In Twitter, salience emerges from the frequent and perceivable communicative practice of ascribing relevance by retweeting. From such a user-centered perspective, the formerly blurred genesis of relevance becomes clearly identifiable, and even distinctively traceable, through the analysis of retweet frequencies. Consequentially, “the corpus processes drive the analysis, and linguistic patterns based around what emerges as frequent or salient need to be accounted for.” (Baker & Levon 2015: 222) Salience then is not only perceivable on the front-end – which is the case for salient retweets but not for the salience of frequently uttered words or phrases – but also becomes detectable by frequency analysis. In Twitter, two objects of analysis might be distinguished when detecting retweet frequencies:

1. the most retweeted tweet and
2. the most retweeted account.[12]

[12] In their study of hashtags in political communication, Barash and Kelly correspondingly focus on “Peakedness, which measures the broad appeal and salience of a contagious phenomenon”(Barash & Kelly 2012: 5).

Assuming that “the most retweeted accounts represent key agents of the protest discourse” (Dang-Anh & Eble 2013: 2; cf. Wilson & Dunn 2011) one should identify these key agents as well as their postings.

4. Twitter

Twitter is, by far, the most popular microblogging platform. The central linguistic unit of analysis is the *posting*, called *tweet* in everyday language. However, the term *tweet* was coined by Twitter Inc. for marketing purposes. Thus, we use the term *posting* as a synonym due to its unrelatedness to specific social media platforms:

> We introduced the category posting as a basic element to capture CMC micro- and macrostructures. A posting is defined as a content unit that is being sent to the server ‘en bloc’. Postings can usually be recognized by their formal structure, even if they have different forms and structures across CMC genres. This facilitates the automatic segmentation and annotation of CMC micro- and macrostructures. (Beißwenger et al. 2012: 5)

With its restriction to 140 characters per posting and its specific distribution features, Twitter is especially useful for time-sensitive digital communication during dynamic events such as street protests. At the time of writing, postings can have four distinct operative in-text features: hashtags, @-mentions, retweets and hyperlinks.

The role Twitter plays in street protests ranges from supportive to constitutive. With the help of Twitter, protestors and observers are able to organize and coordinate protest, inform themselves about what is going on in the streets, mobilize and navigate to relevant places (Dang-Anh & Eble 2013), evaluate situations and political actions, support each other by expressing solidarity, insult or monitor political enemies, build interpersonal relationships and strengthen ties amongst themselves, inform themselves about the course of the protests, celebrate or regret the outcomes or course of a protest event, comment on the protest and so forth (cf. Penney & Dadas 2014). These and many more communicative practices are predominantly performed in Twitter by using four operators: hashtags, retweets, @-mentions and hyperlinks.

4.1 The Operativity of Signs in Digital Communication

The notion of *operativity* refers to characteristics of digital communication platforms and denotes non-human operations that are instructed by human beings through the use of distinct operative characters or buttons. It stems from the concept of the (auto-) operativity of script. Grube (2005) distinguishes between referential script that is assigned to the "writing systems of our narrative and reasoning textual forms" (Grube 2005: 81), operative script that "underlies the cultural technique of written calculation" (Grube 2005: 81) and auto-operative script, "in which single signs of a [notational] system aren't manipulated [by manual operations like written calculation] but automatically processed" (Grube 2005: 82). Operativity in digital communication means that, through the use of operative signs – operators –, authors instruct the machine to execute operations (Grube 2005: 82).[13] These operations include: hyperlinking, including the change of the font color and underlining, sending user notifications, republishing postings and many more.

In Twitter, the executable operations are processed by the practices of hashtagging, retweeting, @-mentioning and hyperlinking whereas the computational operation coincides with the corresponding communicative acts, e.g. citing somebody by retweeting her posting (cf. Thimm et al. 2011). As for our purpose, we will concentrate on the first three functionalities initiated by the usage of the operators '*RT*'[14] or the retweet button, '*#*', and '*@*' and give a brief description of their operativity, as these operators "provide important parameters for quantitative evaluation and qualitative interpretation" (Klemm & Michel 2014: 95).[15]

4.2 Hashtags

Hashtags are used to structure discourses and discourse fragments in Twitter and thus enhance the visibility of tweets (Page 2012). People contribute to a specific topic by using a specific hashtag as "'inline' metadata" that makes topically related tweets searchable and findable (Zappavigna 2011: 791). Furthermore, "Twitter users frequently create idiosyncratic hashtags to add a layer of meaning to a word or phrase" (Dayter 2015: 6). In both respects, hashtags contextualise

[13] The reason for the success of operative digital communication media is its instructional relation between humans and machines: by the rather *simple* use of operators, agents delegate *complex* tasks to the machine and its (black-boxed) algorithmic code. Media linguistics should take such infrastructural processes that lie below the perceptional surface of communication media into consideration. Computer-mediated communication is an everyday business of delegating complex computational processes by simple instructive practices, such as pressing buttons or using operators, without detailed knowledge of these computational processes. An analogue logic is stated by Bishop Berkeley in 1707 for mathematic operations: "The rules may be practiced by men who neither attend to, nor perhaps know the principles … and as any ordinary man may solve divers numerical questions by the vulgar operations of arithmetic, which he performs and applies without knowing the reason of them: Even so… you may operate, compute and solve problems thereby, not only without an actual attention to, or an actual knowledge of the grounds of method." (qtd. in Krämer 1991: 123-124).

[14] The practice of manually adding 'RT' to tweets is not an automated operation but using the retweet button is. Paßmann & Gerlitz point out that the practice of retweeting by typing "RT" and potentially editing or commenting on an initial posting differs from retweeting by using the retweet button (Paßmann & Gerlitz 2014). In the meantime, Twitter has introduced a new function that allows users to add another 140 characters to a button-initiated retweet (cf. Parkinson 2015).

[15] Original: "Möglich machen dies neuartige Vernetzungsstrukturen über technische Operatoren wie RT, # oder @ – die zugleich für die Forschung wichtige Parameter für die quantitative Auswertung wie qualitative Interpretation bereitstellen." (Klemm & Michel 2014: 95)

utterances and thus function as contextualisation cues (Dang-Anh et al. 2013a).
The former specification and thus the selection of messages can be further differentiated by the use of subordinating hashtags. The Twitter discourse of *#PEGIDA*, for example, does not only relate to the anti-Islamic protests in the city of Dresden, where the protests emerged, but also to those that happened in other cities throughout Germany. Further differentiation of discourse was thus done by using other place-related hashtags like *#MAGIDA* for anti-Islamic protest in Magdeburg or *#FRAGIDA* for those in Frankfurt am Main, just to name two of the widely dispersed wave of anti-Islamic protests in Germany in the winter of 2014/15. Counter-protest was often tagged with the prefix "no-", e.g. #NOPEGIDA, #NOMAGIDA or #NOFRAGIDA.

4.3 Retweets

Retweets are used to redistribute and disseminate others' postings, to comment on them, to publicly agree with them, to cite people's utterances, thus making certain tweets and Twitter accounts relevant (boyd et al. 2010). Whereas one measure of the relevance of accounts is the number of followers (Paßmann 2014), another is the amount of retweets of an account's postings. This widens the range of a particular posting, not only at one point in time but, additionally, each and every time a posting is redistributed by retweeting. Thus, the amount of retweets of a protest-related message throughout the course of a protest event is an important measure of relevance whereas the salience of postings and accounts is, as stated above, determined by frequency analysis.

4.4 @-mentions

The @-operator is used to directly address (@-adressing) accounts or to mention (@-mentioning) them within the text of a posting. Depending on the clients' preferences, the addressee might receive a message about his being addressed and thus takes note of it. This may lead to a multiple-turn interaction, though, interactions on Twitter seldom consist of more than an initial tweet-response-structure (cf. Honeycutt & Herring 2009). In protests, @-addressing is often used towards relevant accounts, measured by the quantitative means mentioned above.

5. Three-Step Method

For the sensible combination of quantitative and qualitative analyses, we propose a three-step process of collecting, reducing and interpreting[16] data. In our case, we combine corpus linguistic frequency analysis with sequence analysis that has its origins in ethnomethodological conversation analysis (cf. Bergmann 1981).

1. *Data collection.* It is crucial to identify significant terms and entities as well as an appropriate time span for the segment of discourse that is to be studied. On this basis, the corpus can be compiled accounting for the discourse structuring practices of the agents themselves.[17]

[16] It is vital to note though, that interpretation comes into play at every stage of the research process.
[17] However, there are limitations on collecting data from social media platforms. It is important to note that any selection of identifying linguistic and temporal criteria for data collection involves omitting other data. In social media platforms, access to data is restricted in many ways (cf. boyd & Crawford 2012, Puschmann & Burgess 2014). Therefore, it often is very hard, if not impossible, to determine a basic population for Twitter or Facebook. Additionally, users of social media, and this is especially true for

2. *Data reduction.* Distinct salient agents and utterances can be detected by frequency analysis. Any corpus linguistic method ascertaining frequencies of words, phrases, n-grams and significant co-occurrences etc. might be suitable, depending on your research interest. In social media, textual data is linked with metadata that allows for nexuses of linguistic phenomena such as time, authors, connections between authors, connections between postings, profile pictures, trans- and intermedial links and so forth. For our purpose of linguistic frequency and sequence analysis, the nexus of text, author and time is relevant.
3. *Data interpretation.* Qualitative analysis must be performed with regard to the sequentiality and situatedness of the communicative occurrences. Bringing interactions into sequence reveals interactive phenomena with respect to their characteristics of being jointly constructed. As an act of reflexive contextualisation, hermeneutic qualitative analyses draw on the cotext, the situational and transsituational contexts and background knowledge of the researchers that is methodically substantiated based on the intersubjective insights from hermeneutic-interpretative negotiations.[18]

6. An Exemplary Analysis of a Twitter Dataset on #Fragida

What is #Pegida and #Fragida?

Germany faced a wave of right-wing anti-Islamic protests under the label of *PEGIDA* (*Patriotic Europeans against the Islamisation of the occident*) in winter 2014/2015 (Daphi et al. 2015). The protests originated in Dresden in October 2014 and had several subsidiaries in other German cities. The abbreviation *PEGIDA* was mostly altered with the city's or region's names, e.g. *KÖGIDA* in Cologne (Köln), *BRAGIDA* in Braunschweig, *MAGIDA* in Magdeburg or *BAGIDA* in Bavaria, just to name a few. Soon after the first assemblies of *PEGIDA* in Dresden, counter-protests were established under the label of '*NOPEGIDA*.[19]

6.1 Step 1: Hashtags – Collecting Data Based on Discourse-Relevant Hashtags

The corpus[20] data was collected from the Twitter-Stream-API[21] from 30th January 2015, 0:00:00h to 4th February, 23:59:59h. Every posting that contained selected queries[22] and hashtags identifying the discourse

Twitter, only represent a "very particular sub-set" (boyd & Crawford 2012: 669).

[18] Sequence analysis in digital media, however, is different from conversation analysis of talk-in-interaction due to its *disrupted* sequentiality, e.g. in Twitter, turn sequences are not linear and distinctly distributed from account to account. Beißwenger and Storrer (2008: 301) note: "due to the technical (not pragmatic) sequencing of participant submissions […], there is a larger margin for interpretation (or speculation) in the modelling of CMC data than in the modelling of data from (oral/face-to-face) conversations."

[19] Whereas *NOPEGIDA* is a loose label, the two main actors organising counter-protests in Dresden were *Dresden für alle* (Dresden for everybody) and *Dresden nazifrei* (Dresden free of Nazis). However, for both labels of NOPEGIDA (2015) and NOFRAGIDA (2015) from Frankfurt, there are Facebook groups with several thousand followers.

[20] Unfortunately, Twitter prohibits making the corpus publicly available for further research (cf. Puschmann & Burgess 2014). Although this contradicts our research attitude, we defer to these restrictions due to legal considerations.

[21] For more details on collecting data from Twitter cf. Gaffney & Puschmann (2014).

[22] Search terms were: BAGIDA, BOGIDA, BRAGIDA, DAGIDA, DUEGIDA, DÜGIDA, DUIGIDA, FRAGIDA, HOYGIDA, KAGIDA, KOEGIDA, KÖGIDA, LEGIDA,

on *PEGIDA* and its subsidiaries was collected. Data collection and frequency analysis were performed with the tool *CorpusExplorer*.[23] Data cleansing included eliminating duplicates and spam, and filtering for German tweets. Further data processing comprised lemmatising and tokenising. The corpus consists of 50,633 postings sent by a total of 14,950 accounts and containing 1.06 million tokens of 55,834 types.

We detected an unexpectedly high amount of tweets in the context of the *PEGIDA* protests in Vienna. Qualitative reviewing of these tweets revealed a high ratio of postings from news agencies and media accounts that were not directly related to the street protests. Therefore, we decided to disregard the corresponding Vienna tweets. Hence, through this corpus-hermeneutic analytical step, quantitative analysis could be reassessed by qualitative means circularly.

6.2 Step 2: Retweets – Reducing Data Based on Quantitative Frequency Analysis of Retweets

The most retweeted tweet was posted by the account @Polizei_Ffm on 2nd February, 2015 at 7:55 pm:

8[24]	2015-02-02 19:55	Here the official attendance numbers from tonight: #Pegida/ #Fragida: 85 #nofragida: 1200 #meinfrankfurt #Hauptwache.

It is the "official account of the police Frankfurt am Main during operations" as stated in the profile.[25] In May 2015, the account approved by Twitter had about 20,000 followers. During the data collection period, it was retweeted 94 times and faved[26] 85 times, as

Figure 1: Screenshot of the most retweeted tweet.[27]

displayed below the text in the screenshot (Figure 1). The posting consists of four lines that are separated by line breaks. The first line "Here the official attendance numbers from tonight:" announces the attendance numbers of the protests that occurred on 2nd February, 2015 in the city of Frankfurt am Main. The second line announces the numbers of the attendees of the right-wing anti-Islamic protest. The hashtags *#PEGIDA* and *#FRAGIDA* are used to indicate the reference object of 85 protest attendees. Whereas

MAGIDA, MUEGIDA, MÜGIDA, NUEGIDA, NÜGIDA, OGIDA, PEGIDA, ROGIDA.

[23] CorpusExplorer was programmed by one of the authors, Jan Oliver Rüdiger. It is available as open source software on www.corpusexplorer.de.

[24] This is one of several tweets analyzed in the sequence analysis. The columns from left to right contain: the postings' index, the posting time, the posting text. The whole list of postings is attached in the appendix.

[25] Cf. Polizei_Ffm (2015). As Paßmann (2015: 156) remarks, the white tick in a blue circle (cf. Figure), which marks an account as a *verified account* being identity-proofed by Twitter, indicates a *top-down* assessment by the platform corporation Twitter Inc. whereas assessments like the follower count are *bottom-up*.

[26] For the practice of faving cf. Paßmann & Gerlitz (2014).

[27] This screenshot has been edited: the profile pictures of the retweeting account have been eliminated for anonymization.

#PEGIDA functions as a collective label for the right-wing anti-Islamic protests that were going on throughout Germany at the time, *#FRAGIDA* marks the local reference to the city of Frankfurt. For the opposition, the counter-protests are labelled as *#NOFRAGIDA* and the Frankfurt police counts 1200 attendees in the third line of the posting. The last line contains two hashtags, *#meinfrankfurt*, under which city-related tweets are posted continually and *#Hauptwache* that refers to a building and square in the city.

The high amount of retweets indicates that Twitter users ascribe relevance to this utterance and thus salience that not only becomes visible by frequency analysis but also by the retweet index under the text. As an administrative authority, the police was in operation in the scope of the anti-Islamic protests and the counter-protests in Frankfurt on 2nd February, 2015. Not only because of this prominent societal role of the account but also due to the use of the protest-related hashtags of both political opponents *FRAGIDA* and *NOFRAGIDA*, the tweet might have gained high visibility amongst protestors and observers. Additionally, it thematises the topic of protest attendance numbers, which is another reason for the relevance of the posting. However, from decontextualized analysis of the single posting, it is hard to judge the reason for its ascribed relevance. Thus, it is necessary to contextualize the posting further by looking at the negotiations that the posting evokes and that it is embedded in.

6.3 Step 3: @-mentions – Interpreting Data Based on Sequence Analysis of @-interactions

Sequence analysis in Twitter necessitates putting postings reconstructively and selectively in a reasonable, coherent order. For this purpose and following the reduction of the second step, our analysis focuses

a) on the postings of the most retweeted account (@Polizei_Ffm) that are topically related to the most retweeted tweet (8) and
b) on the interactions that are technically (3,4,5) or topically (9) linked to the these postings (cf. Table 1 in the Appendix).

The first posting regarding the attendance numbers was posted by @Polizei_Ffm at 16:50, 20 minutes after the start time of the counter-protestors' assembly.[28] It states that more than 300 people had yet attended protests at Hauptwache, a square in Frankfurt where the assemblies took place and that is shown in the photo posted with the tweet.

1	2015-02-02 16:50	More than 300 attendees are at the #Hauptwache already in #meinFrankfurt #nofragida #fragida [URL photo]

Whereas in this posting @Polizei_Ffm does not distinguish between the protests of *FRAGIDA* and the counter-protests *NOFRAGIDA*, it does so in further postings. In the course of the protests, @Polizei_Ffm reports on the attendance numbers three more times, including both protests, *FRAGIDA* and *NOFRAGIDA*.

2	2015-02-02 17:37	Intermediate status attendance numbers: #Pegida/ #fragida: 50 #nofragida: 620 #meinfrankfurt #Hauptwache [URL photo]
7	2015-02-02 18:42	Intermediate status attendance numbers: #Pegida/ #fragida: 60 #nofragida: 1180 #meinfrankfurt #Hauptwache [URL photo]

[28] Cf. NOFRAGIDA (2015).

8	2015-02-02 19:55	Here the official attendance numbers from tonight: #Pegida/ #Fragida: 85 #nofragida: 1200 #meinfrankfurt #Hauptwache.

In postings 2 and 7, @Polizei_Ffm chooses an identical wording of the posting with changing attendance numbers. "Intermediate status" here marks a snapshot in the course of the protests and presupposes the anticipation of a rise in attendees. In posting 8, the most retweeted tweet, "official attendance numbers" are reported. In contrast to postings 2 and 7, posting 8 constitutes a final statement regarding the attendance numbers. "Official" here *claims* to state a reliable fact and, as such, interpretational sovereignty over the situation, based on the institutional character of @Polizei_Ffm as an administrative agent. Such a turn harnesses the institutional asymmetry between authorities and non-administrative agents for the (re-) production of an asymmetry of knowledge (Rintel et al. 2013). Consequentially, the assessment of the protest situation is adopted by news media representatives, such as the sender of the following posting, RTL Hessen:

9	2015-02-02 20:14	The @Polizei_Ffm has counted: 85 #Pegida-demonstrants opposed 1200 counter-demonstrators.

It alleges an evidencing process of counting although @Polizei_Ffm has not stated such a process in posting 8.

Looking at the interactions occurring around the theme of attendance numbers, negotiation processes occur:

2	2015-02-02 17:37	Intermediate status attendance numbers: #Pegida/ #fragida: 50 #nofragida: 620 #meinfrankfurt #Hauptwache [URL photo]
3	2015-02-02 17:40	. @Polizei_Ffm The #DankePolizei cannot count: 15 Fragida on the spot .....
4	2015-02-02 17:43	@Polizei_Ffm Where are those 50 hiding? #nofragida
5	2015-02-02 17:45	.@Polizei_Ffm Are the other 47 in the Katharinenkirche just now?
6	2015-02-02 17:53	. @Polizei_Ffm Sorry, the #Pegida number was too high, obviously. Currently, the colleagues have counted 30. Thanks for the pointer.

In postings 3-5, as responses to the initial posting 2 by @Polizei_Ffm, disagreement is expressed. In 3, the ability to count – as stated above: an evidencing process – is doubted. The hashtag *#DankePolizei* (*'thank you, police'*) is conventionally used ironically for postings reporting negatively on police, e.g. cases of police violence. As such, it marks the critical stance towards police authorities. Furthermore, the posting offers an alternative second assessment on attendance numbers as an adjustment of the first assessment (Pomerantz 1984). Postings 4 and 5 ask ironic questions on the whereabouts of *FRAGIDA* attendees. This presupposes an alternate perception on-site and, as such, the posters' physical presence. In posting 6, @Polizei_Ffm revises their first estimation of 50 *FRAGIDA* attendees and downgrades the number to 30. Commenced with an excuse, the police concede misjudgment of attendance numbers. "Obviously" here refers to the congruent commentaries of the respondents on the divergent on-site perceptions. Only in this posting, does @Polizei_Ffm refer to the act of counting as an evidencing practice to emphasize their adjustment of attendance numbers and finally thank the respondents for their allusions.

6.4 Discussion

The analysis provided is exemplary and thus abridged. However, it provides initial insights on the major importance of attendance numbers estimations and reportings for the assessment of political protest. First assessments here were stated by police authorities and controversially negotiated. Raymond & Heritage (2006) identify "three features of assessment sequences" that "are especially relevant for such negotiations" (Raymond & Heritage 2006: 684) and can be applied to protest communication in microblogs:

1) "in order to offer assessments of states of affairs, and so in order to agree or disagree, parties must have some access to them."
2) "speakers rank their access to whatever is being evaluated"
3) "offering a first assessment carries an implied claim that the speaker has primary rights to evaluate the matter assessed" (Raymond & Heritage 2006: 684).

For the police's postings, all three aspects pertain. Holding the monopoly on the legitimate use of force, the police is responsible for the peaceful execution of demonstrations. As such, they assign opposing parties to places and separate them. Thus, they have privileged physical access to protest sites (cf. 1 and 2) thus a structural asymmetry of knowledge (Günthner & Luckmann 1995) is inherent for political protest. As the interactions reveal, physical access and thus the ability to visually perceive protest sites is a precondition for assessments on attendance numbers. However, the respondents (postings 3-5) implicitly claim to somehow have visible access. The asymmetry regarding access is, for the time being, reversed by the police's comment on having (re-)counted the attendance numbers. By posting the "official" attendance numbers in the most retweeted posting, however, asymmetry is restored.

Additionally, asymmetry is also established by the mediality of Twitter: "highly visible users determine what gets amplified and what does not. Twitter's reality is one of asymmetric visibility." (Fuchs 2014: 192). Thus, social relations established by and through medial features within a historical process of social and communicative practices (of relating to each other, e.g. by following someone) predetermine asymmetric relations that themselves reproduce asymmetries amongst social media users.

7. Conclusion

In alignment with the operative mediality of Twitter, we have proposed a simple three-step method in order to heuristically focus from big (amounts of) data to salient practices of communication that can be qualitatively analysed. Firstly, for *collecting* data, it is crucial to identify pertinent terms and entities as well as an appropriate time span for the segment of discourse that is to be studied. For our exemplary Twitter analysis, we have exploratively identified hashtags and words that mark the *PEGIDA* discourse. Secondly, for *reducing* data, distinct salient agents and utterances can be detected by frequency analysis. We selected retweets that indicate the relevance-making practices of the involved agents. Finally, for *interpreting* data, qualitative analysis must be performed with regard to the sequentiality and situatedness of the communicative occurrences. In our case, we have chosen sequential analysis as a means for (re-)constructing the interactional negotiations of meaningful contributions, which are initiated via @-mentions, to protest discourse. However, any qualitative hermeneutic-interpretative method might be applied to salient utterances. Our focus was to describe how corpus linguistics can be helpful as a heuristic tool for qualitative analysis of digital media discourse in media linguistic inquiries.

The internet renders visible a "normative interactive social order" (Thielmann 2012: 101), i.e. it reproduces an asymmetrical relation between administrative and non-administrative agents that negotiate political issues. Asymmetry, in the analysed case, is explicated by the police displaying the reported numbers of demonstration participants as 'official' and is amplified by numerous retweets. Thus, there is both an asymmetry of reputation, regarding the follower counts, and a difference between administrative and non-administrative agents as well as an asymmetry of knowledge which then is negotiated. By assessing and negotiating the number of protest participants with the use of @-mentions, the interlocutors not only evaluate political protest events but make the topic of participants' numbers relevant for protest discourse. Such assessments and negotiation processes – in what Thielmann refers to as accountable social media, i.e. attributable, quantifiable, and visible social media (Thielmann 2012: 100) and by the communicative practices performed in social media platforms – as practices of making-topics-relevant constitute countable as well as assignable indicators of relevance and, as such, salience. Hence, salient communicative practices in digital media are traceable and thus become key data for the analysis of social processes such as protest events. However, it is vital to note that these kinds of analyses focus on communicative practices in the scope of protest events – nothing more, nothing less. In order to achieve deeper levels of understanding and contextualization of communicative practices in digital media, it might be advisable to complement future research designs with additional methods, particularly with ethnographic approaches (cf. Androutsopoulos 2008).

References

Androutsopoulos, J. (2008). Potentials and Limitations of Discourse-Centered Online Ethnography. *Language@Internet*, 5.

Baker, P. & Levon, E. (2015). Picking the Right Cherries? A Comparison of Corpus-Based and Qualitative Analyses of News Articles about Masculinity. *Discourse & Communication*, 9.2, pp. 221-236.

Barash, V. & Kelly, J. (2012). Salience vs. Commitment: Dynamics of Political Hashtags in Russian Twitter. Retrieved from http://ssrn.com/abstract=2034506.

Basic Law for the Federal Republic of Germany, Deutscher Bundestag Art.1 (2012).

Beißwenger, M. et al. (2012). *DeRiK: A German Reference Corpus of Computer-Mediated Communication*. Paper presented at the Digital Humanities 2012 Conference.

Beißwenger, M. & Storrer, A. (2008). Corpora in Computer-Mediated Communication. In A. Lüdeling & M. Kytö (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft = Handbooks of Linguistics and Communication Science: HSK 29.1. Corpus Linguistics: An International Handbook* (pp. 292-309). Berlin, New York: Mouton de Gruyter.

Bergmann, J. (1981). Ethnomethodolgische Konversationsanalyse. In P. Schröder & H. Steger (Eds.), *Dialogforschung. Jahrbuch 1980 des Instituts für deutsche Sprache* (pp. 9-51). Düsseldorf: Pädagogischer Verlag Schwann.

boyd, d. & Crawford, K. (2012). CRITICAL QUESTIONS FOR BIG DATA: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society*, 15.5, pp. 662-679.

boyd, d. et al. (2010). Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. In *Proceedings of HICSS-43* (pp. 1-10).

Bubenhofer, N. (2013). Quantitativ informierte qualitative Diskursanalyse: Korpuslinguistische Zugänge zu Einzeltexten und Serien. In K. S. Roth & C. Spiegel (Eds.), *Angewandte Diskurslinguistik. Felder, Probleme, Perspektiven* (pp. 109-134). Berlin: Akademie Verlag.

Cameron, D. & Panović, I. (2014). *Working with Written Discourse*. Los Angeles: SAGE.

Chafe, W. L. & Nichols, J. (Eds.) (1986). *Evidentiality: The Linguistic Encoding of Epistemology*. Norwood: Ablex.

Dang-Anh, M. & Eble, M. (2013). "Watch out, organize, inform yourself!": Tracing the Dynamics of Twitter Discourse on Anti-Nazi Street Protests. *Selected Papers of Internet Research, IR 14. Resistance + Appropriation*. Retrieved from http://spir.aoir.org/index.php/spir/article/view/729.

Dang-Anh, M. et al. (2013a). Kontextualisierung durch Hashtags. Die Mediatisierung des politischen Sprachgebrauchs im Internet. In H. Diekmannshenke & T. Niehr (Eds.), *Öffentliche Wörter: Analysen zum öffentlich-medialen Sprachgebrauch* (pp. 137-159). Stuttgart: ibidem.

Dang-Anh, M. et al. (2013b). Die Macht der Algorithmen – Selektive Distribution in Twitter. In M. Emmer & I. Stapf (Eds.), *Authentizität in der computervermittelten Kommunikation* (pp. 74-87). Weinheim: Juventa.

Daphi, P. et al. (2015). *Protestforschung am Limit: Eine soziologische Annäherung an Pegida*. Berlin. Retrieved from http://www.wzb.eu/sites/default/files/u6/pegida-report_berlin_2015.pdf.

Dayter, D. (2015). Small Stories and Extended Narratives on Twitter. *Discourse, Context & Media*. Retrieved from http://www.science direct.com/science/article/pii/S2211695815000185.

Entman, R. M. (1993). Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, 43.4, pp. 51-58.

Felder, E. (2012). Pragma-semiotische Textarbeit und der hermeneutische Nutzen von Korpusanalysen für die linguistische Mediendiskursanalyse. In E. Felder et al. (Eds.), *Korpuspragmatik: Thematische Korpora als Basis diskurslinguistischer Analysen* (pp. 116-174). Berlin / Boston: de Gruyter.

Fuchs, C. (2014). *Social Media: A Critical Introduction*. Los Angeles et al.: SAGE.

Gaffney, D. & Puschmann, C. (2014). Data Collection on Twitter. In K. Weller et al. (Eds.), *Twitter and Society* (pp. 55-67). New York: Peter Lang.

Gerbaudo, P. (2012). *Tweets and the Streets: Social Media and Contemporary Activism*. London, New York: Pluto Press.

Gillespie, T. (2010). The Politics of 'Platforms'. *New Media & Society*, 12.3, pp. 347-364.

Grube, G. (2005). Autooperative Schrift – und eine Kritik der Hypertexttheorie. In S. Krämer (Ed.), *Schrift: Kulturtechnik zwischen Auge, Hand und Maschine* (pp. 81-114). München: Fink.

Günthner, S. & Luckmann, T. (1995). Asymmetries of Knowledge in Intercultural Communication: The Relevance of Cultural Repertoires of Communicative Genres. Retrieved from http://kops.uni-konstanz.de/handle/123456789/3694.

Halavais, A. (2014). Structure of Twitter: Social and Technical. In K. Weller et al. (Eds.), *Twitter and Society* (pp. 29-41). New York: Peter Lang.

Honeycutt, C. & Herring, S. C. (2009). Beyond Microblogging: Conversation and Collaboration via Twitter. In R. H. Sprague (Ed.), *Proceedings of the 42nd Annual Hawai'i International Conference on System Sciences. 5-8 January, 2009, Waikoloa, Big Island, Hawaii* (pp. 1-10). Los Alamitos, CA: IEEE Computer Society Press.

Klein, J. (2014). Sätze in der Politik – Strukur. Salienz. Resonanz. In J. Klein (Ed.), *Grundlagen der Politolinguistik: Ausgewählte Aufsätze* (pp. 115-126). Berlin: Frank & Timme.

Klemm, M. & Michel, S. (2014). Big Data – Big Problems? Zur Kombination qualitativer und quantitativer Methoden bei der Erforschung politischer Social-Media-Kommunikation. In H. Ortner et al. (Eds.), *Datenflut und Informationskanäle* (pp. 83-98). Innsbruck: Innsbruck University Press.

Krämer, S. (1991). Zur Begründung des Infinitesimalkalküls durch Leibniz. *Philosophia Naturalis*, 28.2, 117-146.

Linell, P. (1982). *The Written Language Bias in Linguistics*. Linköping: University of Linköping, Dept. of Communication Studies.

Martín Rojo, L. (2014a). Occupy: The Spatial Dynamics of Discourse in Global Protest Movements. *Journal of Language and Politics*, 13.4, pp. 583-598.

Martín Rojo, L. (2014b). Taking over the Square: The Role of Linguistic Practices in Contesting Urban Spaces. *Journal of Language and Politics*, 13.4, pp. 623-652.

McEnery, T. & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge, New York: CUP.

Neumayer, C. & Valtysson, B. (2013). Tweet against Nazis? Twitter, Power, and Networked Publics in Anti-Fascist Protests. *MedieKultur*, 29.55, pp. 3-20.

NOFRAGIDA. (2015). Facebook page of NOFRAGIDA. Retrieved from https://www.facebook.com/NOFRAGIDA1.

NOPEGIDA. (2015). Facebook page of NOPEGIDA. Retrieved from https://www.facebook.com/nopegida.

O'Halloran, K. (2010). How to Use Corpus Linguistics in the Study of Media Discourse. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 563-577). London, New York: Routledge.

O'Keeffe, A. (2012). Media and Discourse Analysis. In J. P. Gee & M. Handford (Eds.), *The Routledge Handbook of Discourse Analysis* (pp. 441-454). London, New York: Routledge.

Page, R. E. (2012). The Linguistics of Self-Branding and Micro-Celebrity in Twitter: The Role of

Hashtags. *Discourse & Communication*, 6.2, pp. 181-201.

Parkinson, H. J. (2015). Digging into Twitter Tunnels: New 'Quote Tweet' Function Launches. Retrieved from http://www.theguardian.com/technology/2015/apr/07/twitter-new-quote-tweet-function-tunnels.

Paßmann, J. (2014). From Mind to Document and Back Again: Zur Reflexivität von Social-Media-Daten. In R. Reichert (Ed.), *Big Data: Analysen zum digitalen Wandel von Wissen, Macht und Ökonomie* (pp. 259-285). Bielefeld: Transcript.

Paßmann, J. (2015). Bewertungssysteme: Medienpraktiken im Umgang mit Fremden. *Zeitschrift für Literaturwissenschaft und Linguistik*, 45.177, pp. 141-166.

Paßmann, J. & Gerlitz, C. (2014). 'Good' platform – political reasons for 'bad' platform-data: Zur sozio-technischen Geschichte der Plattformaktivitäten Fav Retweet und Like. *Mediale Kontrolle unter Beobachtung*, 3.3. Retrieved from http://www.medialekontrolle.de/wp-content/uploads/2014/09/Passmann-Johannes-Gerlitz-Carolin-2014-03-01.pdf.

Penney, J. & Dadas, C. (2014). (Re)Tweeting in the Service of Protest: Digital Composition and Circulation in the Occupy Wall Street Movement. *New Media & Society*, 16.1, pp. 74-90.

Polizei_Ffm. (2015). Twitter Profile of Polizei_Ffm. Retrieved from https://twitter.com/Polizei_Ffm.

Pomerantz, A. (1984). Agreeing and Disagreeing with Assessments: Some Features of Preferred/Dispreferred Turn Shapes. In J. M. Atkinson & J. Heritage (Eds.), *Structures of Social Action. Studies in Conversational Analysis* (pp. 57-101). Cambridge: CUP.

Puschmann, C. & Burgess, J. (2014). The Politics of Twitter Data. In K. Weller et al. (Eds.), *Twitter and Society* (pp. 43-54). New York: Peter Lang.

Raymond, G. & Heritage, J. (2006). The Epistemics of Social Relations: Owning Grandchildren. *Language in Society*, 35.05, pp. 677-705.

Rintel, S. et al. (2013). Knowledge and Asymmetries in Action: Proceedings of the 2012 Conference of the Australasian Institute of Ethnomethodology and Conversation Analysis. *Australian Journal of Communication*, 40.2, pp. 1-8.

Steinberg, R. L. (2014). The Occupy Assembly: Discursive Experiments in Direct Democracy. *Journal of Language and Politics*, 13.4, pp. 702-731.

Thielmann, T. (2012). Taking Into Account: Harold Garfinkels Beitrag für eine Theorie sozialer Medien. *Zeitschrift für Medienwissenschaft*, 6.1, pp. 85-102.

Thimm, C. et al. (2011). Diskurssystem Twitter: Semiotische und handlungstheoretische Perspektiven. In M. Anastasiadis & C. Thimm (Eds.), *Bonner Beiträge zur Medienwissenschaft Vol. 10. Social Media: Theorie und Praxis digitaler Sozialität* (pp. 265-285). Frankfurt am Main: Peter Lang.

Thimm, C. et al. (2014). Mediatized Politics – Structures and Strategies of Discursive Participation and Online Deliberation on Twitter. In A. Hepp & F. Krotz (Eds.), *Mediatized Worlds. Culture and Society in a Media Age* (pp. 253-270). Basingstoke: Palgrave Macmillan.

Tufekci, Z. & Wilson, C. (2012). Social Media and the Decision to Participate in Political Protest: Observations From Tahrir Square. *Journal of Communication*, 62.2, pp. 363-379.

Wilson, C., & Dunn, A. (2011). Digital Media in the Egyptian Revolution: Descriptive Analysis from the Tahrir Data Sets. *International Journal of Communication*, 5, 1248-1272.

Wulff, C. (2010). Speech to mark the Twentieth Anniversary of German Unity. Retrieved from http://www.bundespraesident.de/SharedDocs/Reden/DE/Christian-Wulff/Uebersetzte Reden/2010/101003-Deutsche-Einheit-englisch.html;jsessionid=BF9909033012DE06F3D564F0C1DAAF9F.2_cid293?nn=2748000.

Zappavigna, M. (2011). Ambient Affiliation: A Linguistic Perspective on Twitter. *New Media & Society*, 13.5, pp. 788-806.

Appendix

#	*time*	*account /time of referred posting*	*original text*	*translation*
1	2015-02-02 16:50	Polizei_Ffm	Bereits über 300 Teilnehmer/-innen sind bereits an der #Hauptwache in #meinFrankfurt #nofragida #fragida	More than 300 attendants are at the #Hauptwache already in #meinFrankfurt #nofragida #fragida
2	2015-02-02 17:37	Polizei_Ffm	Zwischenstand Teilnehmerzahlen: #Pegida/ #fragida: 50 #nofragida: 620 #meinfrankfurt #Hauptwache . 	Intermediate status attendance numbers: #Pegida/ #fragida: 50 #nofragida: 620 #meinfrankfurt #Hauptwache [photo]
3	2015-02-02 17:40	@ 17:37	. @Polizei_Ffm Fie #DankePolizei kann nicht zählen: 15 Fragida am Platz	. @Polizei_Ffm The #DankePolizei cannot count: 15 Fragida on the spot .....
4	2015-02-02 17:43	@ 17:37	@Polizei_Ffm Wo haben sich denn die 50 versteckt? #nofragida	@Polizei_Ffm Where are those 50 hiding? #nofragida
5	2015-02-02 17:45	@ 17:37	.@Polizei_Ffm Sind die anderen 47 gerade in der Katharinenkirche?	.@Polizei_Ffm Are the other 47 in the Katharinenkirche just now?
6	2015-02-02 17:53	Polizei_Ffm	. @Polizei_Ffm Sorry, die #Pegida Zahl war offensichtlich zu hoch. Aktuell haben die Kollegen 30 gezählt. Danke für den Hinweis.	. @Polizei_Ffm Sorry, the #Pegida number was too high, obviously. Currently, the colleagues have counted 30. Thanks for the pointer.

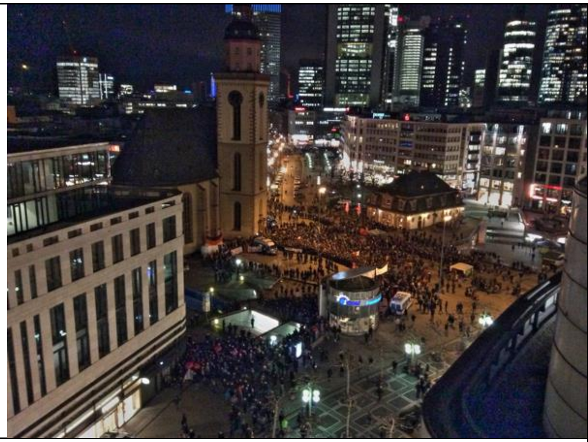
7	2015-02-02 18:42	Polizei_Ffm	Zwischenstand Teilnehmerzahlen: #Pegida/ #fragida: 60 #nofragida: 1180 #meinfrankfurt #Hauptwache . 	Intermediate status attendance numbers: #Pegida/ #fragida: 60 #nofragida: 1180 #meinfrankfurt #Hauptwache [photo]
8	2015-02-02 19:55	Polizei_Ffm	Hier die offiziellen Teilnehmerzahlen des heutigen Abends: #Pegida/ #Fragida: 85 #nofragida: 1200 #meinfrankfurt #Hauptwache.	Here the official attendance numbers from tonight: #Pegida/ #Fragida: 85 #nofragida: 1200 #meinfrankfurt #Hauptwache.
9	2015-02-02 20:14	RTL Hessen	Die @Polizei_Ffm hat durchgezählt: 85 #Pegida-Demonstranten standen 1200 Gegendemonstranten gegenüber.	The @Polizei_Ffm has counted: 85 #Pegida-demonstrants opposed 1200 counter-demonstrators.

Table 1: Postings for sequence analysis