RiCL Research in Corpus Linguistics

How Trump tweets: A comparative analysis of tweets by US politicians

Ulrike Schneider
Johannes Gutenberg-Universität Mainz / Germany

Abstract – This paper analyses tweets sent from Donald Trump's *Twitter* account @realDonaldTrump and contextualises them by contrasting them with several genres (i.e. political and 'average' *Twitter*, blogs, expressive writing, novels, *The New York Times* and natural speech). Taking common claims about Donald Trump's language as a starting point, the study focusses on commonalities and differences between his tweets and those by other US politicians. Using the sentiment analysis tool *Linguistic Inquiry and Word Count* (LIWC) and a principal component analysis, I examine a newly compiled 1.5-million-word corpus of tweets sent from US politicians' accounts between 2009 and 2018 with a special focus on the question whether Trump's *Twitter* voice has linguistic features commonly associated with informality, *I*-talk, negativity and boasting. The results reveal that all political tweets are grammatically comparatively formal and centre around the topics of achievement, money and power. Trump's tweets stand out, however, because they are both more negative and more positive than the language in other politicians' tweets, i.e. his Twitter voice relies far more strongly on adjectives and emotional language.

Keywords – *Twitter*; political communication; sentiment analysis; social media; corpus linguistics

1. Introduction

Sending a number of tweets per day that is unprecedented for a US president, Trump could be termed the first 'social media president' (see also Wodak 2018: xx). Tweets like (1) and (2) are widely known and have helped making terms like *fake news* part of popular culture. Therefore, interest has grown in analyses of the language of these tweets, which is often taken to be crucially different from that of Trump's predecessors and of other politicians. From a cultural and linguistic perspective, however, the language evidenced in the tweets, which I will refer to as 'Trumpish', might merely be a continuation of on-going changes in media use and political culture.

(1) I refuse to call Megyn Kelly a bimbo, because that would not be politically correct. Instead I will only call her a lightweight reporter! (@realDonaldTrump, January 27, 2016)

(2) The Fake News media is officially out of control. They will do or say anything in order to get attention - never been a time like this! (@realDonaldTrump, May 4, 2017)

The present paper provides novel data in order to empirically validate a number of previous claims concerning Trumpish. It studies ‘Trumpish tweets’ from two perspectives; on the one hand, as an idiolect which has become a “branded individual style” (Sclafani 2018: 23). To this purpose, I will contrast the actual language in the tweets to claims about Trump’s idiolect circulating in the media. On the other hand, this paper examines the tweets as an exemplary (if unusually salient) instance of political social media use by comparing the sentiment and (in)formality of tweets sent from Trump’s account to those sent from other politicians’ accounts. Additionally, it also assesses whether the language of the tweets changed as Trump became more involved in politics. Moments that may have triggered a change in the language of the tweets are the day he declared his candidacy (June 16, 2015), the day he was officially nominated the Republican party’s candidate (July 21, 2016), his election (November 8, 2016) and his inauguration (January 20, 2017).

This paper is structured as follows. Section 2 introduces claims about Trumpish circulating in online media and assesses whether those are confirmed by previous analyses —mostly small-scale studies published on linguists’ and laypersons’ blogs. Section 3 introduces our new 1.5-million-word corpus of tweets by Trump and other contemporary US politicians. It further explains how sentiment analysis with LIWC (Pennebaker *et al.* 2015a) and principal component analysis were combined to analyse the corpus. Sections 4 and 5 provide analyses based on grammatical and semantic features of the tweets, followed by the conclusion in Section 6.

2. CLAIMS ABOUT TRUMPISH

This section provides an overview of some of the most commonly repeated claims about Trump’s language and discusses Trump’s idiolect in the context of changes in twentieth-century political communication.

2.1. Trumpish is simple and informal

The internet is rife with analyses finding that spoken Trumpish is characterised by short sentences and simple words (e.g. Crockett 2016; Frischling 2018). The media lapped up findings such as Shafer's (2015) "Donald Trump talks like a third-grader" and particularly Schumacher and Eskenazi's (2016) results, which have been reported as "most presidential candidates speak at grade 6–8 level" (Spice 2016). Such claims are generally based on tests like the Flesch-Kincaid grade-level test, which measure sentence length and number of syllables per word. Their results have to be taken with a grain of salt, though (see Liberman 2015), firstly, because the tests were designed to measure the complexity of *written* language and, secondly, because they react strongly to the way a transcript is punctuated.[1]

Nevertheless, all studies of this kind agree that the length and complexity of Trump's sentences as well as the complexity of his words are among the lowest, if not the lowest, in speeches given by US presidents and candidates in the 2016 race (Shafer 2015; Schumacher and Eskenazi 2016; Rice 2017; Frischling 2018; see also Ronan and Schneider 2020: 73). However, there appears to be some variation between speeches geared to different audiences. This variation could also be the reason why Vrana and Schneider (2017), Frischling (2018), Björkenstam and Grigonitė (2020: 50) as well as Ronan and Schneider (2020: 72) find that Trump's type-token ratio in interviews, speeches, press conferences and debates is lower than that of previous presidents and presidential contenders, while Rice (2017) finds that Trump's inaugural address had an average type-token ratio when compared to 57 previous ones.

Often, (grammatical) simplicity is linked to informality. This is evident in Hunston's (2017) interpretation of her findings. She compares Trump's and Obama's inaugural speeches and finds that Trump's speech is "grammatically simple," consisting of shorter clauses and fewer verbs than Obama's. She then deduces that with their respective speech styles, Obama positions himself as "the statesperson" and Trump as "the ordinary guy." She thus links grammatical simplicity to a more conversational or

[1] Although Schumacher and Eskenazi (2016) as well as Rice (2017) use more sophisticated tests, the problem with potentially deviating punctuation conventions in the transcriptions pertains, because it appears they did not transcribe speeches themselves.

informal speech style and more complex sentence structures to a more formal style.[2] Moreover, Ahmadian *et al.*'s (2017: 51–52) sentiment and part-of-speech analysis of Republican campaign speeches reports that Trump's speeches are significantly less formal than those of the other candidates (p<0.001) and that Trump uses shorter and more non-standard words. Egbert and Biber's (2020: 36) analysis comes to a similar conclusion: compared to other participants in presidential debates, Trump's contributions show fewer signs of "informational language," like "nouns, nominalizations [and] pre-modifying nouns," which results in a "more colloquial, informal tone." Montgomery (2017: 630) even goes as far as saying that Trump uses a "restricted code," i.e. a style tailored to his white, working-class voter base.

Rice's (2017) analysis of 57 presidents puts these findings into a historical context. Trumpish may be simple and conversational, but it is not an outlier. Instead, it is merely the currently last stage in the development of presidential language. In the twentieth century, a shift has taken place in that written language has incorporated features previously restricted to spoken language (e.g. Lakoff 1982: 240; Kowal and O'Connell 1993). Rice's results show that this shift is also evident in the language of (American) politicians of the second half of the twentieth century whose speeches are more conversational than their predecessors' and thus increasingly convey the "warmth, closeness and vividness" (Lakoff 1982: 242, 256) associated with spoken language. In this way, modern politicians deliberately project a "'normal guy' ethos" (Partington and Taylor 2018: 190).

Atkinson (1984: 165–167) argues that it was television which changed the form of political communication towards a "'low-key' television performing style" (see also Kowal and O'Connell 1993: 177) in which politicians use "being relaxed, naturalness, humor, moderate use of gestures and variable formulation as conveyors of the impression of spontaneity and informality" (Kowal and O'Connell 1993: 177 in reference to Atkinson 1984).

Ott (2017: 59) calls this "the Age of Typography" giving way to "the Age of Television." He holds that the twenty-first century has brought yet another turn in political communication, namely towards the "the Age of Twitter" or more generally "the Age of

[2] That there is a general connection between complexity and formality has been established by Biber (1988), who shows that texts with an "interactive, affective, and involved" purpose among other features have fewer long words and a lower type-token ratio than texts with an informational purpose (Biber 1988: 104–107).

Social Media" (Ott 2017: 66), which represents "a fundamental shift in the dominant mode of communication" (Ott 2017: 59). He argues that "[a]s a mode of communication, Twitter is defined by three key features: simplicity, impulsivity, and incivility" (Ott 2017: 59–60). According to Ott (2017: 61) "*Twitter* is structurally ill equipped to handle complex content," which entails that *Twitter* language should *generally* be grammatically simple and constituted of short words (for a similar claim see also Crystal 2011: 20-21). Ott concludes that *Twitter's* influence on public discourse has been so strong that "the Age of *Twitter* virtually guaranteed the rise of Trump" (Ott 2017: 65).

In summary, these observations lead to the expectation that the language used on successful politicians' *Twitter* accounts should be conversational. Particularly, the accounts of presidents and presidential candidates —public personalities with a budget for public relations staff and advisors— should evidence the proposed *Twitter*-style simplicity and informality. Kreis (2017) provides a first indication that in the case of Trump's account this may be true. She concludes from a Critical Discourse Analysis of a selection of Trump's tweets "that his language is simple and direct" (Kreis 2017: 615). Part of Trump's recipe for success might be that he is better at projecting the desired "normal guy ethos" than his competitors (see Montgomery 2017: 624), as proposed by Clarke and Grieve (2019) after a comprehensive analysis of Trump's tweets:

> Trump's Twitter communication style appears to have shifted depending on his intended audience, specifically becoming more informal and conversational when he was trying to appeal to the Republican base and members of the public who shared his political views, and becoming more formal and informationally dense when he was trying to appeal to the general public. […] This strategy was perhaps especially useful for attracting working-class voters […]. These voters may have preferred Trump's informal, unguarded, and outspoken style compared to his competitors in the Republican primaries. Alternatively, shifting to a more formal style during the general election may have helped Trump attract enough independents and moderates to secure his narrow victory over Clinton. (Clarke and Grieve 2019: 20)

However, these studies neither compare his tweets to those of other politicians nor to typical conversational or informal genres. The analysis in Section 2.1 fills this gap by investigating whether Trump's language in the tweets is indeed characteristic of informal genres with narrative content and whether his language is often less formal than that of other politicians.

2.2. Trumpish is I-*talk*

I is by far the most frequent word in the Switchboard NXT corpus of spoken American English, so it does not come as a surprise that it turns out to be Trump's "favorite word" (Shafer 2015). However, media discussions of Trump's pronoun use tap into the psychological associations. Overuse of first person pronouns (*I*-talk) is linked to extraversion, grandiosity and self-focus (Holtgraves 2010: 95–96; Ahmadian *et al*. 2017: 49–50). It has even been stipulated that there is a link between *I*-talk and a Narcissistic Personality Inventory (see discussion in Ahmadian *et al*. 2017: 50). From this perspective, Ahmadian *et al*.'s (2017: 51) finding that Trump uses significantly more first person pronouns in his speeches than the other Republican candidates in their dataset is highly relevant.

Rice's (2017) diachronic analysis of inaugural speeches once more provides a larger perspective. He shows that, in this genre, first person singular pronouns have declined over time. We would thus expect Trump's use to be low —and it is. There are only four tokens in his inaugural speech. The difference between Trump's pronoun use in his inaugural address versus in other speeches might actually result from the use of discourse markers, such as *believe me* in the latter, less rigidly scripted speeches (Sclafani 2018: 36–37).

The use of first-person plural forms, on the other hand, has undergone the reverse development in inaugural addresses. While these pronouns were rare in the earliest speeches, a cross-over took place in the nineteenth century, so that by the end of the century it was more common for presidents to talk about *we*, *us* and *our* than about themselves in the singular. Since then, this trend has continued (with considerable fluctuation; Rice 2017). In light of this, Trump's 98 instances of *we/us/our* observed by Rice (2017) are not even high. The general rise in first person plural pronouns in inaugural speeches may be explained by Atkinson's (1984: 37) finding that

> assertions which convey positive or boastful evaluations of *our* hopes, *our* activities or *our* achievements stand a very good chance of being endorsed by audiences with a burst of applause. [emphasis in the original]

The assessment of parts-of-speech in Section 4.3 will reveal whether *I*-talk is more characteristic of Trump's tweets than of those of other politicians.

2.3. Trumpish is negative and emotional

The public perception of the Trump campaign is perfectly summarised in the following heading from *The Washington Post*: "Welcome to the next, most negative presidential election of our lives" (Blake 2016). Sentiment analyses of the campaign speeches confirm that Trump uses significantly more words with negative connotations than Clinton and, vice versa, that Clinton uses significantly more words with positive connotations than Trump (Jordan and Pennebaker 2016; Hoffmann 2018: 5–6). Furthermore, in his pre-election tweets, 60 per cent of Trump's most frequently used adjectives were negative (e.g. *crooked*, *sad*), while only 20 per cent of Clinton's top adjectives were negative (i.e. *wrong*, *dangerous*; Crockett 2016). Some of Trump's adjectives in the tweets are part of mocking nicknames, like *Crooked Hillary*, *Shady James Comey* or *Crazy Megyn*, which he uses to "diminish and/or discredit his opponents" (Tyrkkö and Frisk 2020: 121).

On the other hand, Crockett's (2016) list shows that Trump's top four adjectives were actually positive (i.e. *great*, *new*, *big*, *amazing*; Crockett 2016). Furthermore, Jordan and Pennebaker's (2016) sentiment analysis also shows that "[d]uring the primary debates, Trump tended to be relatively positive and upbeat." The authors conclude that the tone of Trump's acceptance speech is actually "uncharacteristically negative and pessimistic" for him (Jordan and Pennebaker 2016). These findings put Trump's alleged negativity into question. They rather suggest that Trumpish is a "discourse of dualities" where "the world [is cast] in simple, dualistic terms" (Jamieson and Taussig 2017: 623, 625). Such antonyms and contrasts are commonly used tools of persuasion in politics, particularly in populist rhetoric (e.g. Atkinson 1984: 37–45; Kreis 2017: 609; Pajnik and Sauer 2018 and the papers therein; Partington and Taylor 2018: 51–62).

Kreis' (2017: 607, 615) Critical Discourse Analysis of 200 messages shows that in his tweets "Trump employs positive self-presentation and negative other-presentation to further his agenda." Furthermore, it has been argued that heavy *Twitter* users are people who need a lot of attention and tweet emotional, particularly negative, content in order to get attention (Ott 2017: 62). Thus the question arises whether *Twitter* fosters the "uncivil" and "degrading" side of Trumpish (Ott 2017: 62) or whether it caters to Trump's propensity for "painting pictures of a black-and-white world" (Oborne and Roberts 2017: xi), i.e. to both his negative and his positive side. Clarke and Grieve (2019: 20–21) argue that

> [c]ritical tweets certainly appear to have been an important part of the campaign's communication strategy. However, […] critical tweets did not dominate his timeline […]. Trump and his team appear to have struck an important balance.

The studies in Sections 4.4 and 5.3 attempt to put these findings into a larger context by assessing whether Trump's language is more negative and emotional than would generally be expected.

2.4. Trumpish is boastful

Trump has been described as "a paragon of grandiosity" (Ahmadian *et al.* 2017: 49) and as a frequent user of amplifiers, emphatics and absolutes (Egbert and Biber 2020: 27; Stange 2020: 94–95). In a keyword analysis based on presidential debates, Egbert and Biber (2020: 28) find that

> Trump focuses a great deal on self-promotion and self-aggrandizing statements. This is not unique to Trump, by any means, but he seems to speak in more grandiose terms and to use more repetition when referring to his accomplishments.

A count of the number of boasts in Republican campaign speeches comes to the same conclusion, namely that Trump uses significantly more boasts than the other candidates (p<0.001, Ahmadian *et al*. 2017: 51). Trump's "self-aggrandising" comments have been termed "extreme braggadocio," reminiscent of "ritual boasting" (Montgomery 2017: 625–626). Section 5.4 will provide a comparative analysis of the frequency of achievement-related words in Trump's and other politician's tweets.

3. Data and Method

3.1. Data

My analyses focus on Donald Trump's tweets, but, as a point of comparison, I also use tweets from a range of other politicians' accounts. The list below details the accounts analysed and provides information on how I sourced them.

- **Donald Trump** Account: @realDonaldTrump, 05/2009–07/2018. Source: *Trump Twitter Archive* (Brown 2018), tweets from 2018 were also sourced using *TwitterCorpusQuery 2.0* (Scherl 2018). *Twitter* does not guarantee that the search output contains every tweet matching the search criteria thus some tweets may be missing.

- **Barack Obama** Accounts: @POTUS, 05/2015–11/2016; @WhiteHouse, 05/2009–11/2016. Source: *Internet Archive* (2017).
- **Hillary Clinton** Account: @hillaryclinton, 06/2013–12/2017. Source: *Trump Twitter Archive* (Brown 2018).
- **Sarah Palin** Account: @sarahpalinusa, 11/2009–12/2017. Source: *Trump Twitter Archive* (Brown 2018).
- **Senators** All US senators' accounts, grouped by party. Highlighted Republicans: @JohnCornyn, @senrobportman, @SenTedCruz; Democrats: @amyklobuchar, @SenatorDurbin, @SenWhitehouse; Independents: @SenAngusKing, @SenSanders; 10/2017–03/2018. Source: *TwitterCorpusQuery 2.0* (Scherl 2018).

When analysing the language used in *Twitter* accounts of public figures, what we are seeing is not necessarily the person's own language, but instead the voice officially promoted as Trump's or Clinton's etc., because often a spokesperson is employed who feeds the *Twitter* account. I decided not to attempt to distinguish between staff tweets and tweets of the office holders as a clear distinction would have been impossible despite the fact that Robinson (2016) provides evidence that, in 2016, only tweets sent from an Android system appear to have been written by Trump, while tweets sent from an iPhone appear to have been written by staff. However, a closer look at the data reveals that tweets on the @realDonaldTrump account have also been sent from a variety of other applications (e.g. desktop applications). Furthermore, the link between the Android phone and Trump as the author only seems to hold for 2016, the year of Robinson's analysis, as in 2017, we also find angry, defamatory and emotional tweets coming from an iPhone and in earlier years neither of the two systems had been used for tweeting. Other politicians, like Senator Rob Portman, state on their *Twitter* page that tweets written by the office holders themselves are signed with initials (@robportman via *Twitter.com*; June 21, 2018). However, for technical reasons, many tweets are cut off after 140 characters, which means the signature was lost. Importantly though, politicians communicating with their electorate through spokespeople and/or by means of scripted messages is the default case in current politics. Tweets sent by staff are authorised by the office holder and contribute to what Sclafani (2018: 23) terms a politician's "publically [sic] recognizable branded individual style." Therefore, mixed authorship is not considered noise, but a realistic representation of the way political messages are communicated.

Nevertheless, I undertook several steps to ensure that each dataset is as genuine as possible. Retweets, i.e. messages merely forwarded through the account in question, were excluded. Secondly, as Trump only started using *Twitter*'s official retweet function in 2016 and has instead creatively used different ways of citing others or forwarding their tweets, I also excluded these with the help of simple search algorithms. Finally, all hyperlinks were removed from the tweets. The final corpus contains around 90,000 tweets comprising 1.5 million words.

3.2. Method

In a first step, the tweets were submitted to automatic sentiment analysis using the software *Linguistic Inquiry and Word Count 2015* (LIWC2015, Pennebaker *et al*. 2015a). The programme recognises parts-of-speech and assigns each word to one or more of about 90 grammatical, semantic and punctuation categories. *Cried*, for instance, would be assigned to the following five categories: 'sadness', 'negative emotion', 'overall affect', 'verbs', and 'past focus' (Pennebaker *et al*. 2015b: 2). On average, LIWC recognises around 86 per cent of words in a text (Pennebaker *et al*. 2015b: 10). For each file it has analysed, LIWC calculates the percentage of words assigned to each category. Totals add up to far more than 100 per cent due to multiple category membership. Additionally, the authors have created four summary variables. A complete list of LIWC categories can be found in Pennebaker *et al*. (2015b: 10–12).

As tweets are very short texts, results would fluctuate greatly (see also Clarke and Grieve 2019: 7). For instance, the LIWC output for the single tweet in (3) would state that 100 per cent of its words are adjectives. This exceptionally high rate is misleading as the tweet only consists of a single word. To remedy this, tweets were grouped so as to represent the language used on an account as a whole. Tweets sent from the accounts of all US senators were grouped by party. Additionally, three high-frequency tweeters from each party as well as both independent senators were selected to be represented individually in order to get a more accurate impression of the range of political tweeting styles. Tweets sent from Trump's account were grouped by year in order to be able to distinguish between his language prior and during his political career.

(3) Remarkable! (@SenJohnMcCain, November 3, 2017)

In order to be able to assess whether all political tweets share properties which characterise them as a genre, the analysis also includes LIWC scores provided by Pennebaker *et al*. (2015b). These are based on large corpora representing the following text types: blogs, expressive writing, novels, natural speech, *The New York Times* and *Twitter*. The list below provides some information about the datasets that these LIWC scores are based on. Please note that each text type is represented by a large number of files and that the LIWC scores reported by Pennebaker *et al*. (2015b: 10–11) are means calculated across the files representing a text type.

- *Blogs*. 37,295 blogs downloaded in their entirety from https://.www.blogger.com in 2004. The blogs are balanced for gender and total roughly 119.5 million words (Schler *et al*. 2006; Pennebaker *et al*. 2015b: 9).
- *Expressive writing*. 6,179 essays produced in experiments where participants were asked to write about emotional topics. 2,510 people participated, representing a variety of demographic groups ranging from children to the elderly. This dataset totals 2.5 million words (Pennebaker *et al.* 2015b: 9).
- *Novels*. A sample of 875 novels sourced from *Project Gutenberg*, written between the early seventeenth century and 2008, comprising a total of 57.5 million words (Pennebaker *et al.* 2015b: 9).
- *Natural speech*. Transcripts of spontaneous conversations between acquaintances, couples and strangers, totalling 2.5 million words (Pennebaker *et al.* 2015b: 9).
- *The New York Times*. A sample of around 35,000 articles published in 2014 on *The New York Times* website, representing a variety of sub-genres like news, editorials and letters to the editor. These total 26 million words (Pennebaker *et al*. 2015b: 9–10).
- *Twitter*. Tweets posted from over 35,000 different accounts. On average, each account is represented by around 660 words. These add up to 23 million words (Pennebaker *et al*. 2015b: 9–10).

Thus, the analysis is based on LIWC scores for thirty datasets. 25 of these are compilations of tweets and five represent other genres.

In a second step, the LIWC output was run through principal component analysis (PCA). The aim of this type of analysis is to find clusters of variables which are correlated because they essentially measure the same trait or concept (Field *et al.* 2012: 770; Levshina 2015: 351). For instance, the number of hedges, tag-questions, auxiliaries, epistemic downtoners and references to psychological and social processes have all been shown to measure how feminine a speech style is (e.g. Newman *et al*. 2008). Therefore, they could emerge as a cluster —a so-called dimension— in PCA. Dimensions are not fixed, but instead emerge independently from each dataset. Some are easily interpretable; others may be more difficult to link to linguistic concepts.

This downscaling from many predictors to few dimensions simplifies the data and provides evidence of its meta structure (see also Levshina 2015: 352). Defining a 'supplementary element' in addition to the other factors —the so-called active elements— renders the results more easily interpretable (Levshina 2015: 354). The supplementary element chosen here is nominal and distinguishes between Trump, the politicians and the other text types. The procedure applied here follows Levshina (2015: 351–361). It is conducted in *R* version 3.4.0 (*R* Development Core Team 2009) using the packages *psych* and *FactoMineR* (Le *et al*. 2008).

Only two subsets of the LIWC variables (grammatical and semantic) were selected for analysis. The punctuation-related categories are unreliable due to the cut-off on some tweets and were therefore excluded, as were factors specifically related to spoken language (e.g. disfluencies). Both the restriction to a subset of the available predictors and the split into two subsets help to alleviate the 'large p small n problem' that the data poses (Field *et al.* 2012: 769).

4. Grammatical Analysis

4.1. Data

LIWC provides 21 grammatical categories. I determined the correlations between these, as factors in PCA need to be correlated to some degree, but not perfectly so. Factors which do not have many correlations with a value between 0.3 and 0.9 should be excluded (Field *et al*. 2012: 770–771, 774; Levshina 2015: 353). In the present data, the categories 'common adjectives', 'comparisons', '*we*' and 'numbers' were only weakly correlated with the other factors. However, as the former three are central to my research questions, only the factor 'numbers', which measures the percentage of words such as *second* or *thousand*, was excluded. A Bartlett Test confirmed that the overall degree of correlation in the data was sufficient (χ^2= 1184.382, df=190, p<0.001). (4) lists the final set of grammatical factors considered in the analysis:

(4) total function words, pronouns, personal pronouns, *I*, *we*, *you*, *she/he*, *they*,[3] impersonal pronouns, articles, prepositions, auxiliary verbs, common adverbs, conjunctions, negations, common verbs, common adjectives, comparisons, interrogatives, quantifiers

Due to the large number of factors, 20 dimensions emerged in the PCA. Not all of these were relevant, however. The optimal number of dimensions for a given dataset can be determined by selecting only those whose eigenvalues exceed one (in this case this translates to the dimension explaining roughly 5% of the variance in the data, see Field *et al.* 2012: 764). In the present case, the eigenvalues of five dimensions exceeded one. Yet the fifth dimension barely exceeded this threshold (eigenvalue = 1.16%) and was therefore excluded as well. In addition, the *FactoMineR* package in *R*, which was used here, does not provide information beyond the fourth dimension; therefore, analyses of dimensions beyond the fourth are not possible.

Figure 1 shows the four dimensions. The percentage given next to the dimension label is the variance explained by the dimension. Note that they could have been paired in any way. The selected pairings make it easy to visually analyse the dimensions. Ellipses are 95 per cent confidence intervals around the centroids (Levshina 2015: 360). These could be interpreted as indicating prototypical Trump tweets and prototypical political tweets. Only data-points outside of the main clouds are labelled to improve legibility.

[3] These categories not only include subject pronouns, but also object pronouns and possessive forms (Pennebaker *et al*. 2015b: 3).

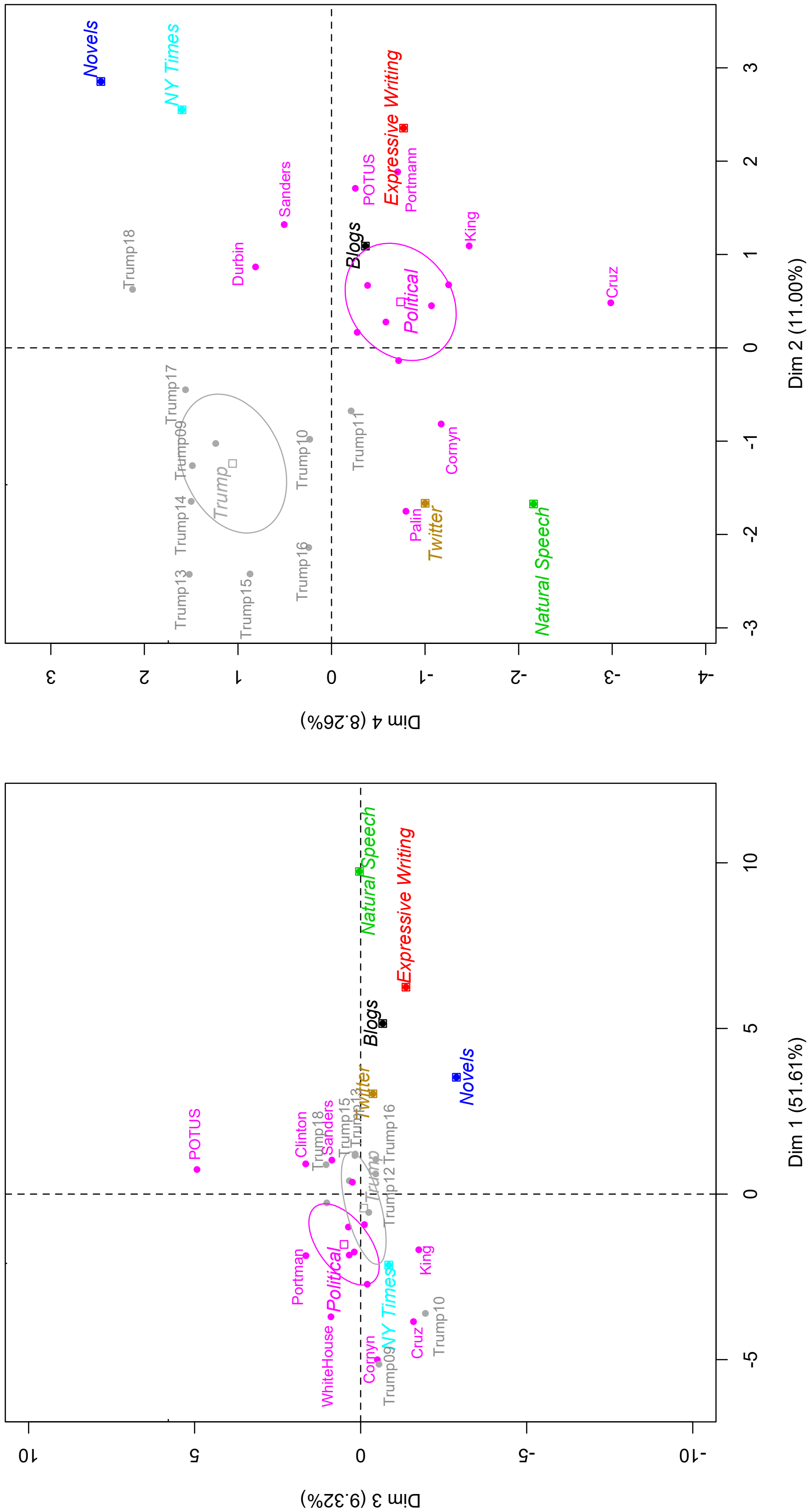

Figure 1: Four main grammatical dimensions identified by the PCA

4.2. Informal language

The question whether Trump's language is less formal or more conversational than other politicians' is addressed by the first dimension (Figure 1, horizontal axis in the left panel). Table 1 shows the factors that are significantly correlated with this dimension.

	Correlation	*p*-value
adverb	0.964	<0.001
pronoun	0.956	<0.001
verb	0.948	<0.001
ppron	0.934	<0.001
function	0.918	<0.001
negate	0.861	<0.001
auxverb	0.858	<0.001
ipron	0.849	<0.001
interrog	0.843	<0.001
conj	0.814	<0.001
I	0.776	<0.001
you	0.655	<0.001
quant	0.607	<0.001
she/he	0.404	<0.05
they	0.370	<0.05
article	-0.399	<0.05
prep	-0.625	<.001

Table 1: Grammatical PCA: Correlations between the quantitative elements and Dimension 1

The underlined items in the list correspond to the parts-of-speech which have been found to determine the degree of formality of a text. Pennebaker *et al*. (2014: 1, 5) state that "greater article and preposition use […] indicat[es] categorical language (i.e., references to complexly organized objects and concepts)" and that "greater use of auxiliary verbs, pronouns, adverbs, conjunctions, and negations […] indicat[es] more dynamic language (i.e., personal narratives)." Additionally, Biber (1988: 102, 107–108) finds that first- and second-person pronouns are strongly associated with texts with "affective, interactional, and generalized content" and that third-person pronouns are "markers of narrative action." Therefore, the first dimension can be interpreted as ranking texts from formal/informationally dense (negative scores) to informal/conversational (positive scores). Thus, the ranking of the genres can be interpreted as follows: from *The New York Times*, via *Twitter*, novels, blogs, expressive writing to natural speech, genres become less formal and more conversational. Interestingly, most political tweets differ considerably from average *Twitter* language in their degree of formality: while average *Twitter* receives a positive (i.e. informal) score, the majority of political tweets are classified as formal language —many are even more formal than *The New York Times*. In fact, 'Political Tweets' is the only supplementary element that is significantly negatively correlated with this dimension ($p<0.05$). Natural speech ($p<0.01$) and expressive writing ($p<0.05$) are positively correlated with it.

While the centroids for Trump and the other political tweeters overlap, indicating that in this regard, their tweets are not fundamentally different, we can still make some interesting observations about Trump's *Twitter* voice. First of all, we can observe a trajectory: in the years between 2009 and 2013, his tweets got ever more informal.[4] After 2012, all years with the exception of 2017 receive a low positive, i.e. moderately informal score. From then on, his tweets are much closer to average *Twitter* than most other politicians'. Interestingly, though, the four political tweeters who also receive positive scores are Palin, Obama (POTUS account), Clinton and Sanders. Particularly the latter three are such high-profile politicians that —if their tweets were not actually formulated for them— we can assume that they had advisors counselling them on how to craft a public image and thus may have been advised to strike a more informal tone.

In summary, we saw that Trump's *Twitter* voice did get more informal/conversational as he developed more of a vested interest in politics. Yet, his *Twitter* voice has never been exceptionally conversational. In fact, it is still more formal than the average tweet, blog or novel. While this may surprise readers who are thinking of tweets such as (5), we have to keep in mind that a large number of the tweets sent from his account rather read like (6), i.e. informal in style but not necessarily grammatically simple.

(5) Big speech tonight in South Carolina - 7:00 P.M. Tremendous crowd! (@realDonaldTrump, February 10, 2016)

(6) How could Jeff Flake, who is setting record low polling numbers in Arizona and was therefore humiliatingly forced out of his own Senate seat without even a fight (and who doesn't have a clue), think about running for office, even a lower one, again? Let's face it, he's a Flake! (@realDonaldTrump, June 7, 2018)

Interestingly, the first dimension also confirms that successful politicians strike a slightly more conversational tone than others. Whether their success results from this style or whether the style was taken on after they became public figures cannot be answered in this type of analysis.

[4] The data-point for 2011 is not labelled. It is located inside the centroid.

4.3. I-*talk*

We saw above that the question has been raised whether Trump's language shows a high degree of self-focus evident in overuse of first-person pronouns. In the present study, *I* does not feature prominently in any of the dimensions. As Table 1 shows, *I* is significantly correlated with Dimension 1, but so are all other personal pronouns except *we*. A look at the LIWC output itself reveals that Trump's *I* use is, in fact, not very high. On average, *I/me/my* make up 2.3 per cent of the words in a dataset and Trump's use is mostly below this. The only years that stand out are 2015 and 2016 where his rates are 2.8 per cent and 2.9 per cent respectively. The only tweeter exceeding this rate is Senator King (2.9%). (7) is an exemplary *I*-tweet sent by Trump in 2016. It reveals that in several instances where *I* could have appeared, it has been dropped (indicated by the added underlines). Overall, the tweets provide no evidence of *I*-talk in Trumpish.

(7) **I** don't know @SamuelLJackson, to best of my knowledge_haven't played golf w/him &_think he does too many TV commercials—boring._Not a fan. (@realDonaldTrump, January 5, 2016)

Crucially, the data shows that grouping singular and plural first-person pronouns together is too simplistic, as they cluster differently in the PCA.[5]

	Correlation	*p*-value
we	0.767	<0.001
compare	0.594	<0.001
quant	0.473	<0.01
adj	0.376	<0.05
she/he	-0.478	<0.01

Table 2: Grammatical PCA: Correlations between the quantitative elements and Dimension 3

Table 2 shows the factors that are significantly correlated with the third dimension. The only account which stands out in Dimension 3 is the POTUS (Obama) account. The amount of *we/us/our* etc. used on this account (4.1%) far exceeds all other tweeters and text types (overall mean: 1.5%), which suggests that Obama portrays a highly confident image (cf. Jordan and Pennebaker 2017). (8) and (9) below show semi-randomly selected tweets by Obama (containing two instances of *we*). (10) and (11) contrast these with tweets containing two instances of *we* sent from Trump's account.

[5] See also the LIWC summary variable 'clout', which measures a speaker's confidence and which is based on findings that *we* and *I* mark opposite ends of the confidence scale (*we* = high confidence, *I* = low confidence; Jordan and Pennebaker 2017; Pennebaker *et al*. 2015b: 6).

(8) **We** could eliminate tuition at every public college and university in America with the $80 billion **we** spend each year on incarcerations. (@POTUS [Obama], July 14, 2015)

(9) 14 months ago, I announced that **we** would begin normalizing relations with Cuba - and **we**'ve already made significant progress. (@POTUS [Obama], February 18, 2016)

(10) When it comes to the future of America's energy needs, **we** will FIND IT, **we** will DREAM IT, and **we** will BUILD IT. #EnergyWeek (@realDonaldTrump, June 29, 2017)

(11) Congress must end chain migration so that **we** can have a system that is SECURITY BASED! **We** need to make AMERICA SAFE! #USA???? (@realDonaldTrump, November 2, 2017)

Note that (8) and (9) with their repeated use of *we* are far more characteristic of the language on Obama's POTUS account than (10) and (11) are of the *we*-use on Trump's account (although (10) and (11) show other characteristic features like all-caps, anxiety and a focus on the future, see Sections 5.2 and 5.3 below).

4.4. Negative and emotional language

A part-of-speech analysis can only provide some indications of the semantics of a text. One such indicator is the frequency of pronouns, as a rhetoric of dualities in terms of positive self-presentation contrasted with negative other-presentation requires the use of pronouns. Furthermore, the use of parts-of-speech like adjectives can be indicative of subjective values being expressed.

The right panel in Figure 1 (see Section 4.1) addresses these issues. In the graph, we see that Trump receives his own quadrant (top left), while most other politicians' tweets are placed in the completely opposite quadrant at the bottom right, indicating that Dimensions 2 and 4 together almost perfectly distinguish the two groups of tweeters. Therefore, the makeup of these dimensions is highly relevant to the distinction between Trump and other political tweeters.

The positive side of Dimension 2 is most strongly correlated with markers of formality (prepositions, articles) as well as with strategies for comparison and quantification (comparatives and quantifiers). This suggests that this side represents written (argumentative) language, which is confirmed by the fact that all written genres receives positive scores, while spoken language and social media (*Twitter*) receives negative scores.[6] In contrast, the negative end of the scale is characterised by adjectives and direct address of interlocutors (second person pronouns), as shown in Table 3.

	Dimension 2			Dimension 4	
	Correlation	*p*-value		Correlation	*p*-value
prep	0.620	<0.001	adj	0.621	<0.001
article	0.516	<0.01	she/he	0.511	<0.01
compare	0.433	<0.05	article	0.500	<0.01
quant	0.410	<0.05	they	0.441	<0.05
conj	0.399	<0.05			
adj	-0.479	<0.01			
you	-0.495	<0.01			

Table 3: Grammatical PCA: Correlations between the quantitative elements and Dimensions 2 and 4

Incidentally, in political tweets, a large number of such direct calls to the audience may be indicative of campaign trail language as Trump's *you* use is particularly high in the years from 2013 to 2016 and the use of *you* in Hillary Clinton's tweets (06/2013–12/2017) is even higher. The examples below show tweets sent by Clinton ((12) and (13)) and Trump ((14)–(17)), each containing several instances of *you*. Note that the Trump examples often seem to follow a template —*Thank you for …*, positive statement, closing slogan— and that they often contain several adjectives.

(12) **You** can knock us down, but **you** can't keep us down. We're always getting up. We're always moving forward. (@hillaryclinton, April 16, 2016)

(13) Whether **you**'re a teacher, an executive, or a world-champion soccer player, **you** deserve equal pay. Red card, GOP. (@hillaryclinton, October 30, 2016)

(14) Wow! This might be my highest # yet! Thank **you** to my opposition- **you** are totally ineffective & have been for years! (@realDonaldTrump, January 22, 2016)

(15) Thank **you** @IvankaTrump for the kind words. I am very proud of the role model **you** are for so many. NH & IA radio ad: [link] (@realDonaldTrump, January 18. 2016)

(16) Thank you to all of the men and women who have served our country. **You** are our true heroes! #ArmedForcesDay (@realDonaldTrump, May 21, 2016)

(17) Thank **you** Bobby Bowden for the intro tonight and **your** support! I hope I can do as well for Florida as **you** have done! (@realDonaldTrump, October 24, 2016)

[6] Social media language, though written, shares features of spoken language (see e.g. Koch and Oesterreicher 2010; Crystal 2011: 20-21).

The adjectives are picked up by Dimension 4, which is the most interesting for othering: those who receive positive scores in this dimension use many adjectives, articles and third person pronouns. As mentioned above, these are parts-of-speech which we would expect to be used to say something negative about others, but, of course, their presence is no guarantee that a tweet is critical, as evident in the examples above, of which only (14) contains a negative adjective. On Dimension 4, Trump once more receives scores which are the opposite of the other political tweeters' scores. His frequent use of adjectives plays a large role: when the data is ranked by adjective use, the ten Trump datasets make up the top third —only Palin and Obama use as many adjectives as he does. The semantic analysis below will reveal whether Trump's adjectives are more often negative, indicating outspoken, critical language or whether they are actually mostly positive.

The use of third-person pronouns is also picked up by Dimension 4. Trump's rate is consistently around the mean or above it. There are a few years which stand out, though. In 2009 and 2010, hardly any *they* occurs in the tweets sent from Trump's account (0.2% and 0.1% of words respectively). (18) shows one of the tweets from 2010 which does contain a third person plural pronoun —it references dunes, not people.

(18) The Dunes here are <u>amazing</u>, and **they**'re how I learned about geomorphology, which is the study of movement landforms. We've had a <u>great</u> trip (@realDonaldTrump, May 27, 2010)

The year 2017 also differs from the rest of the Trump tweets in terms of third-person-plural pronoun use. Tweets from this year contain fewer *she/he* but more *they* than the other years. (19) and (20) are specimens of tweets in which *they* occurs at least twice. Both tweets are negative and antagonistic.

(19) If Republican Senators are <u>unable</u> to pass what **they** are working on now, **they** should immediately REPEAL, and then REPLACE at a <u>later</u> date! (@realDonaldTrump, June 30, 2017)

(20) The <u>Fake</u> News refuses to talk about how <u>Big</u> and how <u>Strong</u> our BASE is. **They** show <u>Fake</u> Polls just like **they** report <u>Fake</u> News. Despite only <u>negative</u> reporting, we are doing well - nobody is going to beat us. MAKE AMERICA <u>GREAT</u> AGAIN! (@realDonaldTrump, December 24, 2017)

The *they* trend continues in 2018, suggesting that, as president, Trump talks about groups rather than individuals. Overall, Dimensions 2 and 4 show that Trumpish is characterised by the use of adjectives as well as by second- and third-person pronouns.

5. Semantic Analysis

5.1. Data

Originally, 50 LIWC factors were included in the semantic dataset, but the strength of the majority of correlations did not exceed 0.3. Therefore, the decision was made to exclude all predictors whose correlation with the other factors did not exceed 0.3 in at least 25 per cent of cases. (21) lists the final set of semantic factors considered in the analysis. A Bartlett Test confirmed that the overall degree of correlation in the data was sufficient (χ^2= 7937.103, df=946, p<0.001).

(21) positive emotion, negative emotion, anxiety, anger, sadness, social processes, friends, female references, male references, cognitive processes, insight, causation, discrepancy, tentative, certainty, differentiation, perception, see, feel, biological processes, body, health, ingestion, drives, affiliation, achievement, power, reward, risk, past focus, present focus, future focus, relativity, motion, space, work, leisure, home, money, death, informal language, swear words, netspeak, assent

Due to the large number of factors, 29 dimensions emerged, nine of which had eigenvalues exceeding 1. However, only the top four will be discussed. This decision was made based on the limited information the package provides after the fourth dimension and, furthermore, because the percentage of variance explained dropped abruptly from over ten per cent to 5.7 per cent after the fourth dimension. Figure 2 shows the resulting dimensions. Please, note again that the dimensions could have been grouped in any way and that I selected groupings which make it easy to inspect the data visually.

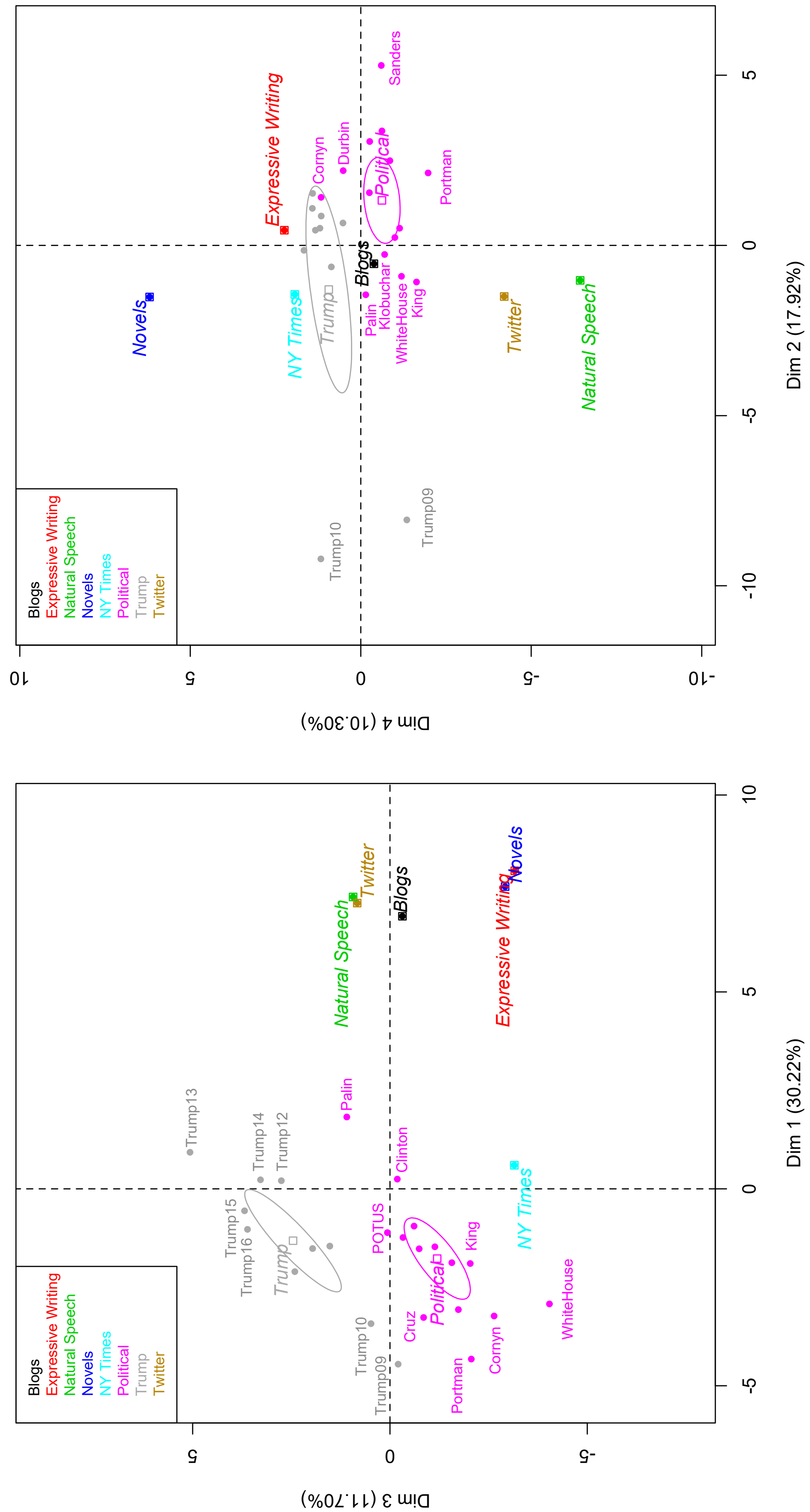

Figure 2: Four main semantic dimensions identified by the PCA

5.2. Informal language

Dimension 4 (Figure 2, vertical in the right panel) can be cautiously interpreted in relation to formality. Its negative end is associated with features of informal, conversational spoken language as well as with informal written language (i.e. netspeak, e.g. *btw*, *lol*; focus on the present, e.g. verbs, *now*; assent, e.g. *OK*, *yes*; see Table 4). However, these features are not contrasted with more formal language on the positive end of the scale. Instead, they are opposed by features which indicate frustration and/or past events as topics —i.e. anxiety, sadness, focus on the past, which are exemplified in (22) and (23).

(22) Jeb Bush never uses his last name on advertising, signage, materials etc. Is he ashamed of the name BUSH? A pretty sad situation. Go Jeb! (@realDonaldTrump, August 24, 2015)

(23) Passing what was once a vibrant manufacturing area in Pennsylvania. So sad! #MakeAmericaGreatAgain (@realDonaldTrump, April 25, 2016)

	Correlation	*p*-value
male	0.632	<0.001
anx	0.478	<0.01
space	0.470	<0.01
sad	0.443	<0.05
focuspast	0.442	<0.05
female	0.424	<0.05
negemo	0.419	<0.05
death	0.401	<0.05
netspeak	-0.632	<0.001
focuspresent	-0.635	<0.001
assent	-0.683	<0.001
informal	-0.690	<0.001

Table 4: Semantic PCA: Correlations between the quantitative elements and Dimension 4

Once more, we see that political tweets are rather unlike average *Twitter* —the latter being rated as far more informal than the former. While most politicians receive moderately negative scores, Trump receives exclusively positive scores (except for 2009), placing him closer to *The New York Times* than to *Twitter* and natural speech. Overall, political tweets —whether written by Trump or others— cluster around blogging language in this case.

5.3. Negative and emotional Language

Dimension 3 answers the question whether Trump's vast amount of adjectives is positive or negative. Crucially, both positive and negative emotions are positively correlated with this dimension (see Table 5), meaning that they are not treated as opposites, but as features characterising the same texts —and those texts are Trump's tweets. Figure 2 (left panel, vertical axis) shows that Trump receives the highest scores of all tweeters and text types, which means that his tweets are typically both more negative and more positive than the other politicians' tweets and, in fact, more so than average *Twitter*.

	Correlation	*p*-value
reward	0.770	<0.001
certain	0.749	<0.001
posemo	0.703	<0.001
negemo	0.657	<0.001
sad	0.542	<0.01
focusfuture	0.499	<0.01
anger	0.493	<0.01
focuspresent	0.456	<0.05
bio	-0.501	<0.01
health	-0.597	<0.001
home	-0.736	<0.001

Table 5: Semantic PCA-Correlations between the quantitative elements and Dimension 3

Trump's tweets actually started out with rather unemotional language (i.e. with a score just below zero), but in the years between 2009 and 2013, the score increased every year, so that all tweets after 2009 are rated as emotional —only Obama (POTUS account) and Palin also receive positive scores, indicating that their tweets are characterised by emotions. The latter two, however, still rank in the region of natural speech and average *Twitter*, while Trump's score far exceeds this. Consequently, the centroids around the means do not overlap on this scale, indicating that the prototypical Trump tweet differs significantly from the prototypical political tweet, which is also confirmed by the fact that Trump's tweets are highly significantly positively correlated with this dimension ($p<0.001$), while the other political tweets are significantly negatively correlated with it ($p<0.01$).

In summary, the language of the tweets sent from the @realDonaldTrump account is characterised by emotions in general and anger and sadness in particular (presumably the word *sad* itself). Trump's tweets from 2009 and 2010 once more receive more moderate scores than those from other years, which is due to them neither being emotional nor showing strong signs of reward and certainty-oriented thinking. The duo of negative and positive emotions is an indicator of dualistic thinking, while the certainty/reward combination indicates that the argumentation may be based on the simplistic notion of "direct causation" (Lakoff 2016).

5.4. Bragging

Finally, the semantic analysis provides information about the topics addressed in the tweets. We will take a look at Dimensions 1 and 2, the two horizontal axes in Figure 2.

Dimension 1 (left panel) in particular provides information about the level of focus on achievements in the data. This dimension sees almost all of the political tweeters on the left side (significantly negatively correlated with the dimension, $p<0.05$), while all other genres receive positive scores. Table 6 shows that the negative end of the scale is characterised by work-related terms (i.e. *money, work*) and by 'drives' (i.e. *achievement, power*).

Dimension 1			Dimension 2		
	Correlation	*p*-value		Correlation	*p*-value
tentat	0.929	<0.001	cause	0.830	<0.001
feel	0.901	<0.001	risk	0.756	<0.001
cogproc	0.853	<0.001	discrep	0.692	<0.001
differ	0.849	<0.001	death	0.634	<0.001
body	0.847	<0.001	anger	0.621	<0.001
insight	0.822	<0.001	money	0.560	<0.01
ingest	0.815	<0.001	drives	0.557	<0.01
focuspast	0.797	<0.001	affiliation	0.522	<0.01
swear	0.701	<0.001	anx	0.481	<0.01
bio	0.672	<0.001	negemo	0.469	<0.01
female	0.668	<0.001	home	0.423	<0.05
informal	0.608	<0.001	social	0.413	<0.05
assent	0.561	<0.01	differ	0.385	<0.05
male	0.530	<0.01	health	0.373	<0.05
anx	0.492	<0.01	cogproc	0.367	<0.05
friend	0.491	<0.01	sad	-0.375	<0.05
motion	0.466	<0.01	focusfuture	-0.555	<0.01
netspeak	0.460	<0.05	relativ	-0.603	<0.001
social	0.422	<0.05	percept	-0.738	<0.001
space	-0.484	<0.01	see	-0.789	<0.001
money	-0.627	<0.001	leisure	-0.796	<0.001
drives	-0.705	<0.001			
achieve	-0.759	<0.001			
work	-0.800	<0.001			
power	-0.886	<0.001			

Table 6: Semantic PCA — Correlations between the quantitative elements and Dimensions 1 and 2

Example (24) below provides an example of a tweet with several expressions referencing money and achievements.

(24) I have not heard any of the pundits or commentators discussing the fact that I spent FAR LESS MONEY on the win than Hillary on the loss! (@realDonaldTrump, December 21, 2016)

The opposite end of the scale is characterised by topics like feelings, thoughts, people and the body. Thus, the dimension distinguishes between a focus on achievements and a focus on feelings. We can conclude that political *Twitter* as a whole is (not totally unexpectedly) characterised by a focus on money, power and achievement. It furthermore shows that this is one of the major differences between political *Twitter* and average *Twitter* (the latter being significantly positively correlated with the dimension, $p<0.05$).

Overall, Trump does not differ from the other political tweeters. Nevertheless, we see a trajectory, which is by now familiar: his tweets from 2009 and 2010 are very much achievement-oriented, but with every year they become less so. The tenor shifts towards feelings, thoughts and people up to a zenith in 2013, after that, he swings back slightly. So, if anything, some of his *Twitter* years stand out for being less focussed on achievements than typical political tweets.

Dimension 2 (Figure 2, right panel, horizontal axis) sheds some further light on the topics of the tweets. It explains the difference between Trump's tweets from 2009 and 2010 and his later tweets. The former are all about the future, perception (particularly seeing) and leisure. This is the case because they contain frequent imperatives on what to watch, see and read in the future, as shown in (25).[7]

(25) Be sure to <u>look</u> for my beautiful wife Melania Trump tonight on QVC at 9 pm ET where she <u>will be</u> debuting her fantastic jewelry collection. (@realDonaldTrump, April 30, 2010)

6. Conclusion

The present analysis has shown that many common assumptions about Trump's language either do not hold up to scrutiny or cannot be generalised to his tweets. Firstly, there were no indications in the data that Trump's tweets are exceptionally informal or conversational in terms of the parts-of-speech which are used. Instead, like all political tweets, they were rather formal in this respect. This means that, while Trump's debate contributions stand out for having fewer prepositions and articles than other participants' as well as more pronouns and adverbs (see e.g. Egbert and Biber 2020), these features are not what characterise his tweets. Secondly, no indication of *I*-talk could be found in Trump's tweets. Thirdly, and most importantly, it turned out that Trump's tweets may be more negative than other political tweets, but that they are also more positive. His *Twitter* voice relies far more strongly on adjectives and emotional language than other political

[7] My emphasis.

Twitter accounts. Finally, it transpired that all political *Twitter* centres around power, work and achievements, and Trump's is no exception.

We also observed a number of changes in Trump's *Twitter* voice. Tweets from 2009 and 2010 are clearly not political tweets —they rather promote the Trump brand, then consisting of beauty pageants, Trump University, books, golf courses, a reality TV show, casinos, hotels and TV appearances. This is in line with Clarke and Grieve's (2019) finding that the style of Trump's tweets from those years is 'advisory', while it turns 'critical' in the following years. Still, right from the start of his tweeting career, we see Trump's voice developing: between 2009 and 2013 Trump's tone becomes increasingly less formal and more emotional. By 2013 he seems to have found 'his voice', which he later moderates a little, but continues to use mostly unchanged until today. This shift also transpires in Clarke and Grieve's (2019) analysis of the tweets. They find a peak in conversational style in 2013 and a shift from the 'critical' to another 'advisory' period around that time. The final set of years which stands out on some scales are 2015 and 2016, which are characterised by emotional language, many adjectives and second-person pronouns. This could be interpreted as Trump's campaign-trail style (compare a peak in campaign trail style —though determined based on different parameters— found for the same period by Clarke and Grieve (2019)).

The study also provides a characterisation of prototypical US political *Twitter*: generally a formal text type with many characteristics of written language, centred on work, achievement, money and power. On each scale, a couple of tweeters stand out, often because they show similar deviations from 'political norm-*Twitter*' as Trump. On several occasions, these are Senator Cornyn and Sarah Palin, both conservative Republicans. Though on other scales, these are Obama (POTUS account), H. Clinton and Senator Sanders, moderate Democrats and Independents with high public visibility.

References

Ahmadian, Sara, Sara Azarshahi and Delroy L. Paulhus. 2017. Explaining Donald Trump via communication style: Grandiosity, informality, and dynamism. *Personality and Individual Differences* 107: 49–53.

Atkinson, Max. 1984. *Our Masters' Voices: The Language and Body Language of Politics*. London: Routledge.

Biber, Douglas. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

Björkenstam, Kristina Nilsson and Gintarė Grigonitė. 2020. I know words, I have the best words. Repetitions, parallelisms, and matters of (in)coherence. In Ulrike Schneider and Matthias Eitelmann eds., 41–61.

Blake, Aaron. 2016. Welcome to the next, most negative presidential election of our lives. *The Washington Post*. https://www.washingtonpost.com/news/the-fix/wp/2016/07/29/clinton-and-trump-accept-their-nominations-by-telling-you-what-you-should-vote-against/?noredirect=on&utm_term=.0faae7f1e872 (6 July, 2018.)

Brown, Brendan. 2018. *Trump Twitter Archive*. http://www.trumptwitterarchive.com/ (10 April, 2018.)

Clarke, Isobelle and Jack Grieve. 2019. Stylistic variation on the Donald Trump *Twitter* account: A linguistic analysis of tweets posted between 2009 and 2018. *PLoS ONE* 14/9: e0222062.

Crockett, Zachary. 2016. What I learned reading 4,000 Trump and Clinton tweets. *Vox*. https://www.vox.com/2016/11/7/13550796/clinton-trump-twitter (12 April, 2018.)

Crystal, David. 2011. *Internet Linguistics. A Student Guide*. London: Routledge.

Egbert, Jesse and Douglas Biber. 2020. 'It's just words, folks. It's just words'. Donald Trump's distinctive linguistic style. In Ulrike Schneider and Matthias Eitelmann eds., 17–40.

Field, Andy, Jeremy Miles and Zoë Field. 2012. *Discovering Statistics Using R*. London: Sage.

Frischling, Bill. 2018. 'Stable genius' – Let's go to the data. *Factbl.og*. https://factba.se/blog/2018/01/08/stable-genius-lets-go-to-the-data/ (12 April, 2018.)

Hoffmann, Thomas. 2018. 'Too many Americans are trapped in fear, violence and poverty': A psychology-informed sentiment analysis of campaign speeches from the 2016 US Presidential Election. *Linguistics Vanguard* 4/1: 1–9.

Holtgraves, Thomas. 2010. Text messaging, personality, and the social context. *Journal of Research in Personality* 45/1: 92–99.

Hunston, Susan. 2017. Talking Trump: Literally speaking. *University of Birmingham*. https://www.birmingham.ac.uk/research/perspective/talking-trump-literally-speaking.aspx (16 April, 2018.)

InternetArchive. 2017. *Obama White House Twitter Archive*. https://archive.org/details/ObamaWhiteHouseTwitterArchive (11 September, 2017.)

Jamieson, Kathleen Hall and Doron Taussig. 2017. Disruption, demonization, deliverance, and norm destruction: The rhetorical signature of Donald J. Trump. *Political Science Quarterly* 132/4: 619–650.

Jordan, Kayla N. and James W. Pennebaker. 2016. Accepting the nomination: A comparison of the speeches of Trump and Clinton. https://wordwatchers.wordpress.com/2016/08/01/accepting-the-nomination-a-comparison-of-the-speeches-of-trump-and-clinton/ (10 April, 2018.)

Jordan, Kayla N. and James W. Pennebaker. 2017. Trump's first State of the Union Address. https://wordwatchers.wordpress.com/2017/03/01/trumps-first-state-of-the-union-address/ (10 April, 2018.)

Koch, Peter and Wulf Oesterreicher. 2010. Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch* 36: 15–43.

Kowal, Sabine and Daniel C. O'Connell.1993. Television rhetoric in an age of secondary orality: Psycholinguistic analyses of the speaking performance of Ronald Reagan.

Georgetown Journal of Languages and Linguistics 1: 174–185. Translated reprint of: Kowal, Sabine and Daniel C. O'Connell. 1993. Fernsehrhetorik im Zeitalter der zweiten Mündlichkeit: Psycholinguistische Analysen des Sprachverhaltens von Ronald Reagan. In Paul Goetsch and Gerd Hurm eds. *Die Rhetorik amerikanischer Präsidenten seit F.D. Roosevelt*. Tübingen: Gunter Narr, 247–260.

Kreis, Ramona. 2017. The 'tweet politics' of President Trump. *Journal of Language and Politics* 16/4: 607–618.

Lakoff, George. 2016. Understanding Trump. https://georgelakoff.com/2016/07/23/understanding-trump-2/ (2 March, 2018.)

Lakoff, Robin. 1982. Some of my favourite writers are literate: The mingling of oral and literate strategies in written communication. In Deborah Tannen ed. *Spoken and Written Language. Exploring Orality and Literacy.* Norwood, NJ: Ablex, 239–260.

Le, Sebastien, Julie Josse and François Husson. 2008. FactoMineR: An *R* Package for Multivariate Analysis. *Journal of Statistical Software* 25/1: 1–18.

Levshina, Natalia. 2015. *How to Do Linguistics with R. Data Explorations and Statistical Analysis*. Amsterdam: John Benjamins.

Liberman, Marc. 2015. More Flesch-Kincaid grade-level nonsense. http://languagelog.ldc.upenn.edu/nll/?p=21847 (6 July, 2018.)

Montgomery, Martin. 2017. Post-truth politics? *Journal of Language and Politics* 16/4: 619–639.

Newman, Matthew L., Carla J. Groom, Lori D. Handelman and James W. Pennebaker. 2008. Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes* 45/3: 211–236.

Oborne, Peter and Tom Roberts. 2017. *How Trump Thinks: His Tweets and the Birth of a New Political Language.* London: Head of Zeus.

Ott, Brian L. 2017. The age of Twitter: Donald J. Trump and the politics of debasement. *Critical Studies in Media Communication* 34/1: 59–68.

Pajnik, Mojca and Birgit Sauer eds. 2018. *Populism and the Web. Communicative Practices of Parties and Movements in Europe.* London: Routledge.

Partington, Alan and Charlotte Taylor. 2018. *The Language of Persuasion in Politics. An Introduction.* London: Routledge.

Pennebaker, James W., Cindy K. Chung, Joey Frazee, Gary M. Lavergne and David I. Beaver. 2014. When small words foretell academic success: The case of college admissions essays. *PLoS ONE* 9/12: 1–10.

Pennebaker, James W., Roger J. Booth, Ryan L. Boyd and Martha E. Francis. 2015a. *Linguistic Inquiry and Word Count: LIWC2015.* Austin, TX: Pennebaker Conglomerates (www.LIWC.net).

Pennebaker, James W., Ryan L. Boyd, Kayla Jordan and Kate Blackbun. 2015b. *The Development and Psychometric Properties of LIWC2015*. Austin, TX: University of Texas at Austin.

R Development Core Team. 2009. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. http://www.R-project.org.

Rice, Justin. 2017. Does Trump really have the best words? https://www.litcharts.com/blog/analitics/does-trump-really-have-the-best-words/ (27 June, 2018.)

Robinson, David. 2016. Text analysis of Trump's tweets confirms he writes only the (angrier) Android half. http://varianceexplained.org/r/trump-tweets/ (10 October, 2017.)

Ronan, Patricia and Gerold Schneider. 2020. A man who was just an incredible man, an incredible man. Age factors and coherence in Donald Trump's spontaneous speech. In Ulrike Schneider and Matthias Eitelmann eds., 62–86.

Scherl, Magdalena. 2018. *TwitterCorpusQuery 2.0*. Mainz.

Schler, Jonathan, Moshe Koppel, Shlomo Argamon and James W. Pennebaker. 2006. Effects of age and gender on blogging. *Proceedings of AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs*. Stanford, CA. https://www.aaai.org/Papers/Symposia/Spring/2006/SS-06-03/SS06-03-039.pdf

Schneider, Ulrike and Matthias Eitelmann eds. 2020. *Linguistic Inquiries into Donald Trump's Language. From 'Fake News' to 'Tremendous Success'*. London: Bloomsbury.

Schumacher, Elliot and Maxine Eskenazi. 2016. *A Readability Analysis of Campaign Speeches from the 2016 US Presidential Campaign.* Pittsburgh, PA Language Technologies Institute, School of Computer Science: Carnegie Mellon University.

Sclafani, Jennifer. 2018. *Talking Donald Trump*. London: Routledge.

Shafer, Jack. 2015. Donald Trump talks like a third-grader. *Politico*. https://www.politico.com/magazine/story/2015/08/donald-trump-talks-like-a-third-grader-121340 (12 April, 2018.)

Spice, Byron. 2016. Most presidential candidates speak at grade 6-8 level. *Carnegie Mellon University News*. https://www.cmu.edu/news/stories/archives/2016/march/speechifying.html (10 April, 2018.)

Stange, Ulrike. 2020. *Very emotional*, *totally conservative*, and *somewhat all over the place*. An analysis of intensifiers in Donald Trump's speech. In Ulrike Schneider and Matthias Eitelmann eds., 87–108.

Tyrkkö, Jukka and Irina Frisk. 2020. *Crooked Hillary*, *Lyin' Ted*, and *Failing New York Times*: Nicknames in Donald Trump's Tweets. In Ulrike Schneider and Matthias Eitelmann eds., 109–129.

Vrana, Leo and Gerold Schneider. 2017. Saying whatever it takes: Creating and analyzing corpora from US presidential debate transcripts. *Extended Abstracts of Corpus Linguistics Conference*. 24–28 July 2017, Birmingham.

Wodak, Ruth. 2018. Preface. From 'hate speech' to 'hate tweets'. In Mojca Pajnik and Birgit Sauer eds., xvii–xxiii.

Corresponding author
Ulrike Schneider
Department of English and Linguistics
Johannes Gutenberg-Universität Mainz
Jakob-Welder-Weg 18
55128 Mainz
Germany
e-mail: ulrike.schneider@uni-mainz.de

received: February 2020
accepted: December 2020