

ABSTRACT

DISSERTATION: Interactions³: Language, Demographics, and Personality; an In-Depth Analysis of German Tweets

STUDENT: Matthias Ludwig Raess

DEGREE: Doctor of Philosophy

COLLEGE: Sciences and Humanities

DATE: May 2018

PAGES: 260

This study examines the interactions between German Twitter users' personalities, the specific features of their tweets, including the emoji density, types of hashtags and their density, and the percentage of various LIWC word categories, such as emotion words (e.g. positive/negative), and how these variables interact with gender. This tripartite analysis in conjunction with questionnaire data and online Twitter data not only advances our understanding of how gendered-language, and thus perceived stereotypes, and personality in conjunction with linguistic cues behaves on Twitter, but also how an out-of-the-lab sample contributes to the generalizability of results and insights gleaned from these analyses. The significance, therefore, not only lies in the combination of research areas (linguistics and psychology) and the analysis approach, but also the fact that Twitter studies focusing on German, incorporating personality measures are far and few between.

The broad research design of this study is quantitative and interdisciplinary in nature, encompassing both linguistics and psychology. Participants, $N = 62$, filled out an online questionnaire providing demographic information and information on their personality traits through the German short version of the Big Five Inventory, the BFI-10 (Gosling, Rentfrow, &

Swann, 2003; Rammstedt & John, 2007; Rammstedt, Kemper, Klein, Beierlein, & Kovaleva, 2012). In addition, participants' tweets, $N = 19,772$, were collected using the Twitter API and then combined with their demographic information, including their Big5 scores. The tweets were then analyzed with the software, Linguistic Inquiry and Word Count (LIWC) (Pennebaker & King, 1999), which made it possible to quantify linguistic features, such as percentage of emotion words, or anger words, for example. In addition, quantitative measures of users' frequencies of hashtags, including a hand-coded hashtag subset, $n = 2,666$, and emojis, including sentiment scores were used for quantitative analyses.

The study furnishes new confirmatory evidence for previous findings regarding significant positive correlations between positive feeling words and extraversion, agreeableness, and neuroticism. A significant positive correlation between neuroticism and anxiety words was also confirmed. In addition, extraversion turned out to be a significant predictor for sentiment scores on Twitter, indicating that extroverts benefit more from being active online. In terms of LIWC categories, gender was a significant predictor for both positive emotion words, and positive feeling words, with females using more in both categories. Gender also turned out to be a significant predictor for anger words, swear words, occupation words, and words related to money, with women using higher percentages in these categories, which contradicts previous research in an English-language context, and contrary to my own expectations.

The study thus offers new insights into the differences in relationship to gender and context-dependent language use, adding support for some of the key arguments. Specifically, female German Twitter users turned out to use language differently compared to previous findings, e.g. lower percentage of words related to tentativiy. In addition, German female Twitter users seem to use the social medium for different, more professional purposes. The study

thus addresses contentious previous findings by adding new information to the perceptions of gender-dichotomous language use, indicating that it does not necessarily follow the same patterns across genres, i.e. different social media, prompting a re-thinking of some previous findings. The statistical significance of the findings allows us to make conservative and careful generalizations to the larger German Twitter user base.

Keywords: LIWC, sociolinguistics, Twitter, German, Big Five, language and gender