BERTAgent: A Novel Tool for Quantifying Linguistic Agency: Development, Validation and an Application in the 2020 U.S. Election Tweets

Keywords: machine learning, linguistic agency, collective action, communication, politics

Extended Abstract

Agency can be defined as the propulsion that orients people toward action and prompts them to achieve their goals (Abele & Wojciszke, 2014; Bandura, 2001), as such it is an important aspect of social dynamics and of exchanges that set people into collective action (e.g., Loehr, 2022). Social media (SM) play an important role in coordination of collective phenomena, such as political processes (e.g., Brady et al., 2017; Goldenberg & Gross, 2020; Schöne et al., 2021). SM also provide voluminous data to study the interplay between linguistic agency and collective action. Yet, we still miss a valid tool able to reliably capture linguistic agency. Existing computational approaches predominantly rely on a dictionary word count methodology that is based on a simple count of words considered to be indicative of agency. The simplicity of this approach, however, results in a number of important shortcomings that are symptomatic for all dictionary based methods. For example, the lack of sensitivity to polysemy (e.g., "this water contains lead" vs. "we can lead them"), negation (e.g., "not motivated" vs. "motivated" if no additional rules are applied), and the lack of nuance in how well words represent a given construct, because -in the dictionary approach- all words have the same weight (e.g., "try" vs. "strive"). Finally, dictionary-based tools are often insensitive to construct directionality; that is, they do not differentiate between words that imply lack of agency and words that are indicative for the presence of agency (e.g., "unmotivated" vs. "aspiring"). Here, we present BERTAgen, a novel tool that alleviates these problems by utilizing a machine learning approach to natural language processing. BERTAgent is a computational language model fine-tuned on carefully selected textual data that were extracted from a lexicographical database and evaluated by human participants (*N*=567) with respect to the level of agency. BERTAgent is sensitive not only to particular words but also to the context they are used in. This work describes BERTAgent development stages and results of five validation studies using a diverse sample of natural language (*N*=7,850; *N*=132; *N*=216; *N*=972; *N*=1,288). Overall, BERTAgent outperforms previous tools in terms of convergent and discriminant validity. The procedure used to develop BERTAgent is generalizable and can be employed to develop tools that will quantify other psychological constructs.

We also demonstrate an example application of BERTAgent in which we assess the prevalence of agentic language in tweets (*N*=495,252) posted by candidates (*N*=1283) in the 180-day time-span prior and post the 2020 U.S. Congressional Elections. This timespan allowed us to examine linguistic behavior of politicians with respect to major expected (e.g., election) and unexpected political events (e.g., 6th January Capitol Attack). In both types of the events politicians are expected to mobilize their followers and accordingly use more agentic language prior to expected and after unexpected events. We base this expectation on the following assumptions. In order to build and maintain strong support, politicians must be able to project an image of effectiveness and capacity to act (agency). To achieve that, they must not only be effective in their actions, but also effectively communicate and transfer their agency to trigger public support (e.g., vote). Accordingly, studies show perceived candidates' agency such as proactivity and need for achievement drive their leadership evaluations, which in turn predict voting behavior (Pillai et al., 2003). Furthermore, elections can be considered an example of collective action, where individuals join their efforts to achieve a common goal (van Zomeren, 2013). Importantly, a sense of agency is among the core predictors of engagement in collective action. Therefore, the efforts to mobilize voters around a certain cause should increase prior to an expected event or spike right after an unexpected event. **Methods.** ***Data collection.*** We identified (*N*=1283) Twitter profiles of Democrat (D, *N*=454) and Republican (R, *N*=416) candidates using Ballotpedia. For each profile we downloaded all available tweets for the period of 180 days before and after the election. This resulted in a dataset containing 495,252 tweets (for detailed breakdown of tweet counts see Figure 2; for daily tweet count timecourse see Figure 3). ***Linguistic features extraction.*** We quantified agency in tweets using BERTAgent. We also controlled for other linguistic features that have been previously found to be important for political messages (e.g., concreteness, readability, valence and emotions). **Results and Conclusions.** Regression analysis revealed that election candidates use more linguistic agency prior to election day (and other expected events), also, in line with our expectations, we found bursts of agentic language after significant unexpected events (see Figures 4 & 5). Strategic and instrumental approach to agency can be attested by sudden and substantial decrease in agency after elections. Moreover, this strategy appears to be effective as higher levels of agency prior to elections are related to successful election outcomes. Overall, BERTAgent offers high precision and accuracy in detecting linguistic agency, allowing for interesting applications in the domain of computational social science.

References

Abele, A. E. & Wojciszke, B. (2014). Communal and agentic content in social cognition: A dual perspective model. Advances in experimental social psychology (pp. 195–255). Elsevier. https://doi.org/10.1016/B978-0-12-800284-1.00004-7

Bandura, A. (2001). Social Cognitive Theory: An Agentic Perspective. Annual Review of Psychology, 52(1), 1–26. https://doi.org/10.1146/annurev.psych.52.1.1

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. Proceedings of the National Academy of Science, 114, 7313–7318. https://doi.org/10.1073/pnas.1618923114

Formanowicz, M., Pietraszkiewicz, A., Roessel, J., Suitner, C., Witkowska, M. & Maass, A. (2021). "Make it Happen!": Verbs as markers of agency increase message effectiveness. Social Psychology, 52(2), 75–89. https://doi.org/10.1027/1864-9335/a000435

Loehr, J. D. (2022). The sense of agency in joint action: An integrative review. Psychonomic Bulletin & Review, 29(4), 1089–1117. https://doi.org/10.3758/s13423-021-02051-3

Pietraszkiewicz, A., Formanowicz, M., Gustafsson Sendén, M., Boyd, R. L., Sikström, S. & Sczesny, S. (2019). The big two dictionaries: Capturing agency and communion in natural language. European Journal of Social Psychology, 49(5), 871–887. https://doi.org/10.1002/ejsp.2561

Pillai, R., Williams, E. A., Lowe, K. B. & Jung, D. I. (2003). Personality, transformational leadership, trust, and the 2000 U.S. presidential vote. The Leadership Quarterly, 14(2), 161–192. https://doi.org/10.1016/S1048-9843(03)00008-0

Schöne, J. P., Parkinson, B. & Goldenberg, A. (2021). Negativity Spreads More than Positivity on Twitter After Both Positive and Negative Political Situations. Affective Science, 2(4), 379–390. https://doi.org/10.1007/s42761-021-00057-7

van Zomeren, M. (2013). Four Core Social-Psychological Motivations to Undertake Collective Action: Four Motivations for Collective Action. Social and Personality Psychology Compass, 7 (6), 378–388. https://doi.org/10.1111/spc3.12031

Zhu, X., Kim, Y. & Park, H. (2020). Do messages spread widely also diffuse fast? Examining the effects of message characteristics on information diffusion. Computers in Human Behavior, 103, 37–47. https://doi.org/10.1016/j.chb.2019.09.006

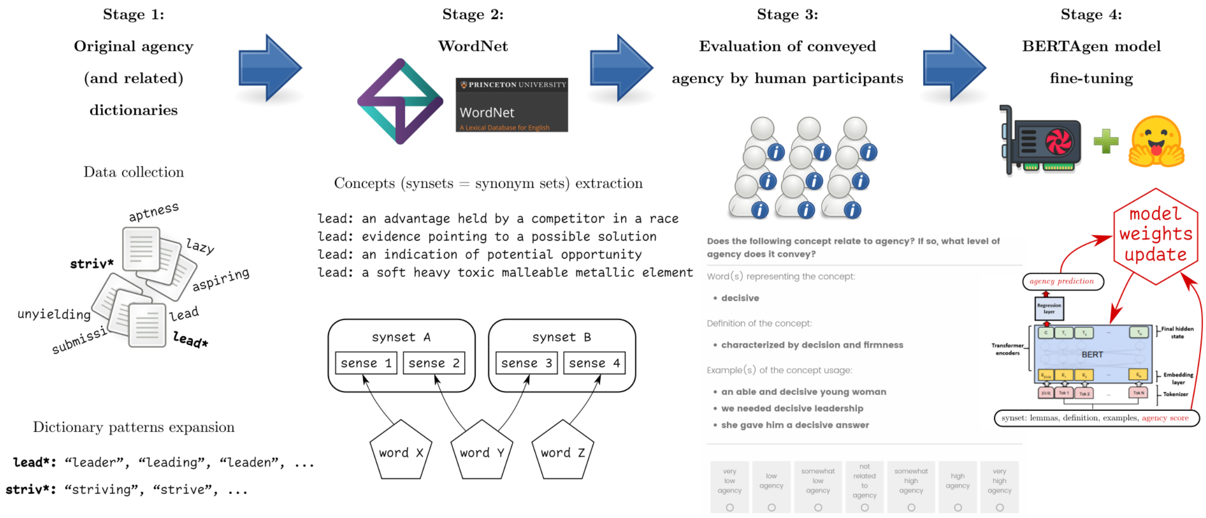

Figure 1. Overview of BERTAgent Development Stages. Stage 1: Selection of agency related word forms and stem patterns from the existing DWC-based tools; Stage 2: Extraction of synsets (synonym sets, akin to concepts) from the WordNet database; Stage 3: Evaluation of concepts by human participants; Stage 4: Language model fine-tuning on the regression task.

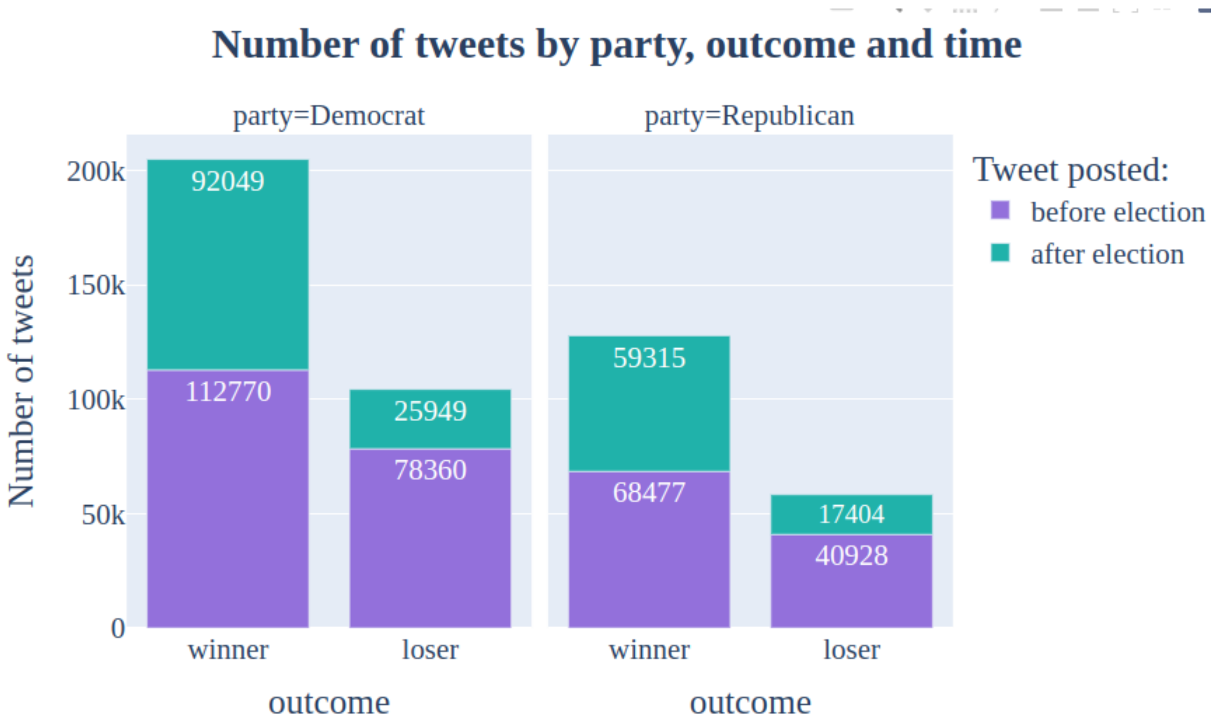

Figure 2. Number of tweets by party, outcome and timing (relative to election day).

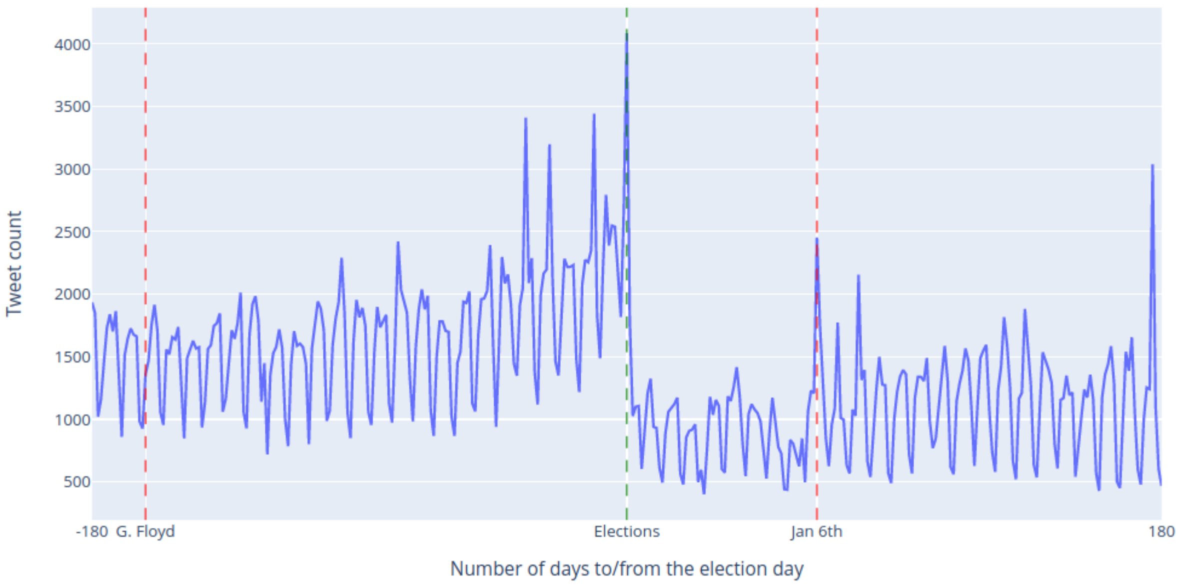

Figure 3. Daily tweets count timecourse.

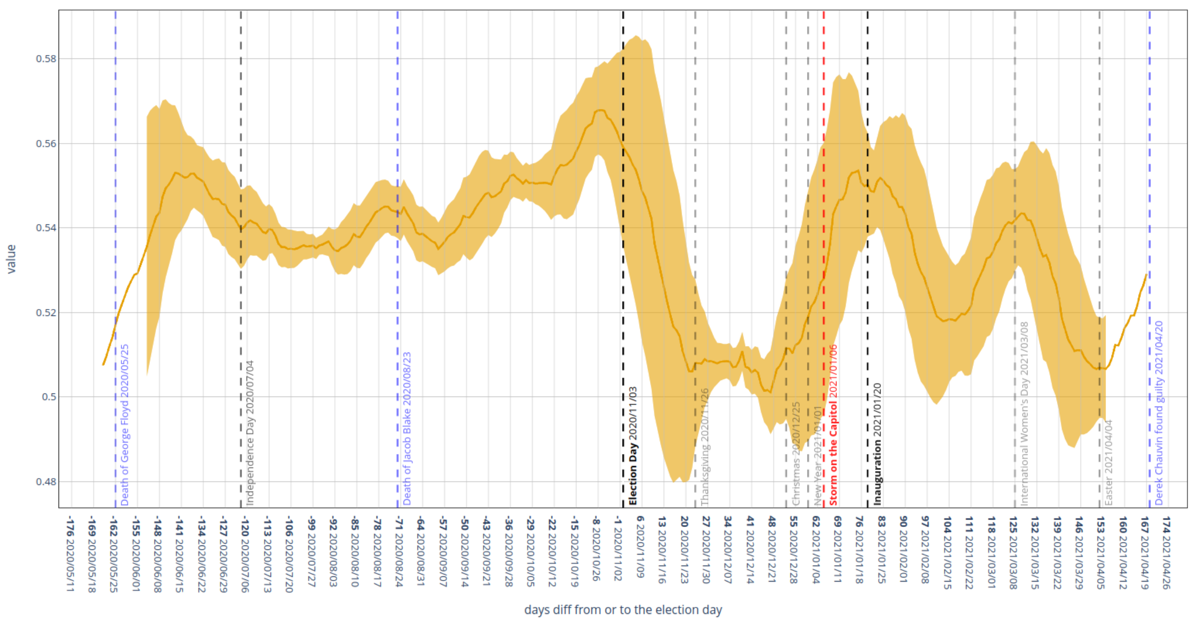

Figure 4. Overall mean level of linguistic agency of tweets 180 days before and after election date. Additional politically significant events include: January 6th United States Capitol attack and G. Floyd death. Data was smoothed using a Gaussian filter (FWHM=14 days).

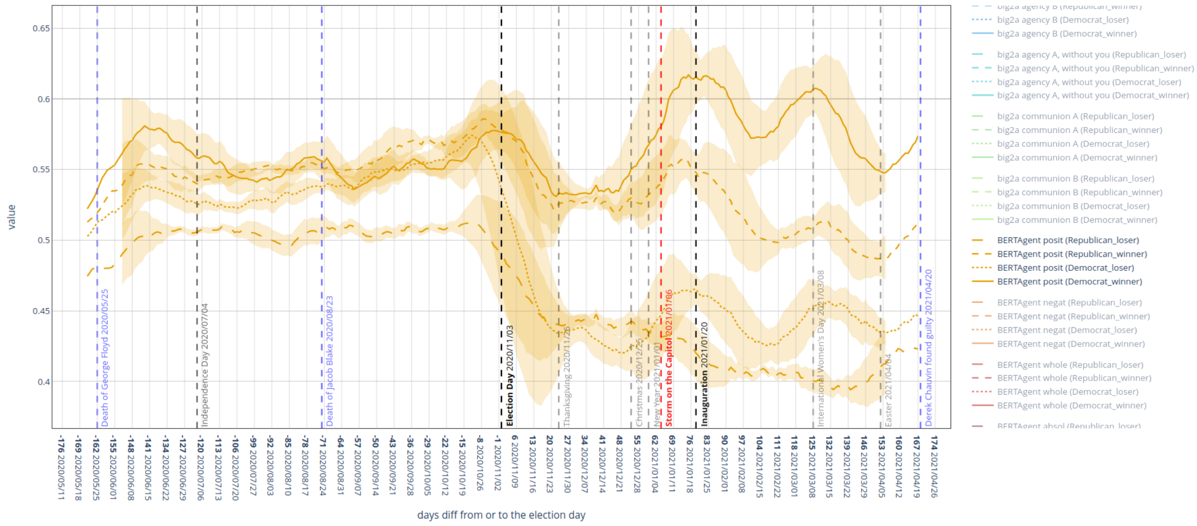

Figure 5. Mean level of linguistic agency of tweets 180 days before and after election date split according to political party (R or D) and election outcome (winner or loser). Additional politically significant events include: January 6th United States Capitol attack and G. Floyd death. Data was smoothed using a Gaussian filter (FWHM=14 days).