How Does Fact-checking Message Intensity Influence Public Engagement and Attitude towards Fact-checking Targets: A Computational Analysis of Facebook Messages

Keywords: misinformation correction, fact-checking effectiveness, language intensity, social media engagement

Extended Abstract

Background

Fact-checking, a systematic validation of claims, myths, and rumors, is effective in correcting misperceptions and spreading true information (Walter et al., 2019). Though it is not a new form of journalism, the public is unfamiliar with the growing fact-checking industry (Robertson et al., 2020). It is necessary for the agencies to disseminate fact-checking messages to a wide audience to correct the misinformed and inoculate the uninformed. Sensationalistic news features such as intense language can be used to attract attention and engagement (Burgers & Graaf, 2013). However, such practice can also backfire and elicit suspicions from the reader and run into the problem of losing credibility (Brandtzaeg et al., 2017). So far, very little research has examined how language intensity, a nuanced linguistic feature reflecting message emotionality and attitudinal extremity (Bowers, 1963), influences message engagement and the public's specific attitude toward fact-checking targets.

Grounded on the information processing model of language intensity (Hamilton & Stewart, 1993), this study investigates (a) how language intensity of fact-checking messages influences the public's engagement, (b) how language intensity influences the public's attitude toward fact-checking targets and fact-checking agencies, and (c) how fact-checked information veracity moderates the effect of language intensity. This study helps us understand how linguistic features may implicitly impact the transmission of fact-checking messages and the public's attitude toward fact-checking targets. This study also has practical implications to guide fact-checking efforts to effectively communicate facts on social media.

Methods

Data collection (Figure 1). This study relies on fact-checking posts (N = 49,270) and comments (N = 525,604) on Facebook between October 2017 and October 2022 created by 10 fact-checking agencies verified with the International Fact-Checking Network (IFCN).

Language intensity. We operationalized language intensity as the sentiment magnitude, representing the strength of emotion provided by the Google Cloud Natural Language Application Programming Interface (API), ranging from 0 to positive infinity.

Engagement with fact-checking messages is measured by all engagement behaviors on Facebook: the number of likes, comments, shares, and 6 emotional reactions such as *Angry*. A composite score of all 9 metrics is created to capture the overall message engagement.

Public's attitude toward fact-checking targets and agencies. We employed entity-targeted sentiment analysis by Google Cloud Natural Language API to (a) identify entities mentioned in both a post and comments attached, (b) manually annotate fact-checking targets (i.e., issues, public figures, government organizations) and fact-checking agencies (e.g., snopes.com), and (c) operationalize the public's attitude as the entity-targeted sentiment score in the comments, ranging from extreme negativity (-1) to extreme positivity (1).

Fact-checked information veracity. We parsed the fact-checking labels from the links attached to the posts and manually categorized 45 labels that have been used more than 5 times into *false* (25 labels; e.g., altered, fake), *half-true* (13; e.g., mixture, outdated), *true* (3; e.g., legit), and *unknown* (4; e.g., unproven, research-in-progress) for further analysis.

Results

First, language intensity was positively associated with overall engagement (Figure 2). Specifically, fact-checking posts with higher language intensity tended to have higher levels of engagement, especially the number of shares, and most emotional reactions (i.e., *Angry*, *Haha*, *Sad*, *Wow*). However, language intensity was negatively associated with *Love* and *Care*. Language intensity can boost the public's engagement and negative emotional reactions to fact-checking messages but inhibit positive emotional expressions.

Second, language intensity was negatively associated with the public's overall attitude expressed in comments and the attitude toward government organizations (e.g., government, CDC, congress, cops), suggesting higher intensity generates more negative attitudinal reactions. However, language intensity was not significantly associated with the attitude toward fact-checking issues (e.g., health, crude oil) or fact-checking agencies (e.g., snopes.com, PolitiFact).

Third, fact-checking agencies used more intense language in reporting fact-checked false information. We observed the negativity bias: language intensity was significantly higher for *false* labels than *half-true* labels and *true* labels in a pairwise t-test (Figure 3).

Lastly, fact-checked information veracity moderated the effect of language intensity on engagement and the public's attitude (Figure 4). The positive effect of language intensity on engagement was significantly higher for *false* labels than for *half-true* and *true* labels; language intensity was negatively associated with engagement for *unknown* labels (b = -0.29, p = .006). In addition, we observed a backfiring effect: language intensity was negatively associated with the public's overall attitude and the attitude toward government organizations for fact-checked *true* information.

Conclusion

This study systematically examined fact-checking label intensity. High-intensity fact-checking messages had more public engagement but elicited more negative public attitudes, especially toward government organizations. Fact-checking agencies used more intense language in reporting *false* information, which in turn obtained higher levels of engagement. On the contrary, higher intensity in fact-checked *true* information received less attention but was associated with a more negative attitude toward government organizations.

References

Bowers, J. W. (1963). Language intensity, social introversion, and attitude change. *Communication Monographs*, *30*(4), 345–352. https://doi.org/10.1080/03637756309375380

Brandtzaeg, P. B., Følstad, A., & Domínguez, M. Á. C. (2017). How Journalists and Social Media Users Perceive Online Fact-Checking and Verification Services. *Journalism Practice*, *12*(9), 1109–1129. https://doi.org/10.1080/17512786.2017.1363657

Burgers, C., & Graaf, A. de. (2013). Language intensity as a sensationalistic news feature: The influence of style on sensationalism perceptions and effects. *Communications*, *38*(2), 167–188. https://doi.org/10.1515/commun-2013-0010

Hamilton, M. A., & Stewart, B. L. (1993). Extending an information processing model of language intensity effects. *Communication Quarterly*, *41*(2), 231–246. https://doi.org/10.1080/01463379309369882

Robertson, C. T., Mourão, R. R., & Thorson, E. (2020). Who Uses Fact-Checking Sites? The Impact of Demographics, Political Antecedents, and Media Use on Fact-Checking Site Awareness, Attitudes, and Behavior. *The International Journal of Press/Politics*, *25*(2), 217–237. https://doi.org/10.1177/1940161219898055

Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2019). Fact-Checking: A Meta-Analysis of What Works and for Whom. *Political Communication*, *37*(3), 350–375. https://doi.org/10.1080/10584609.2019.1668894
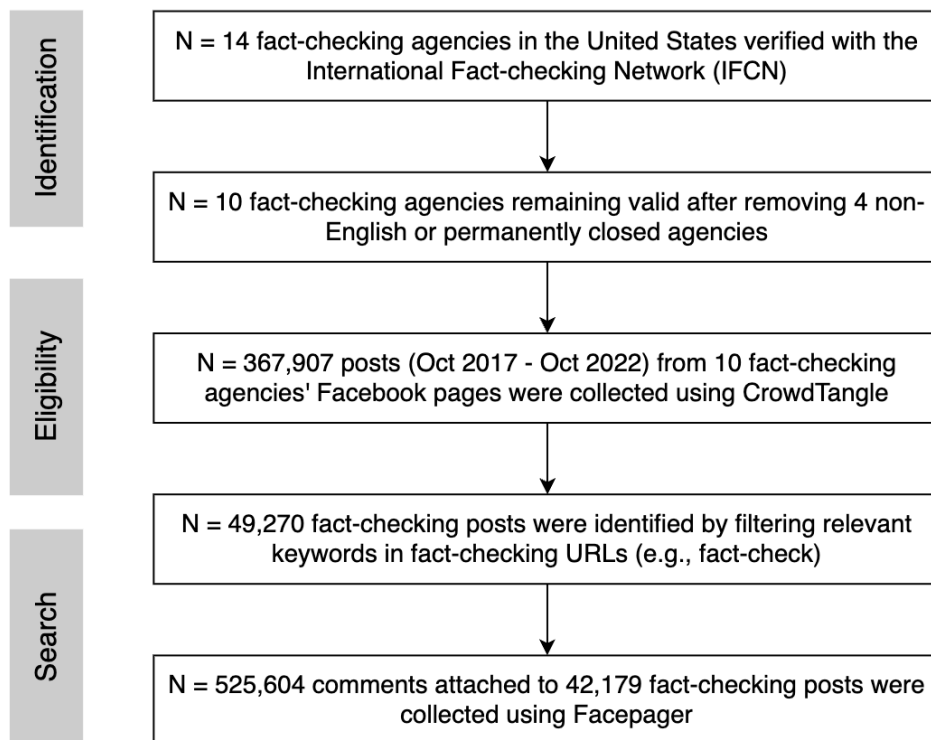
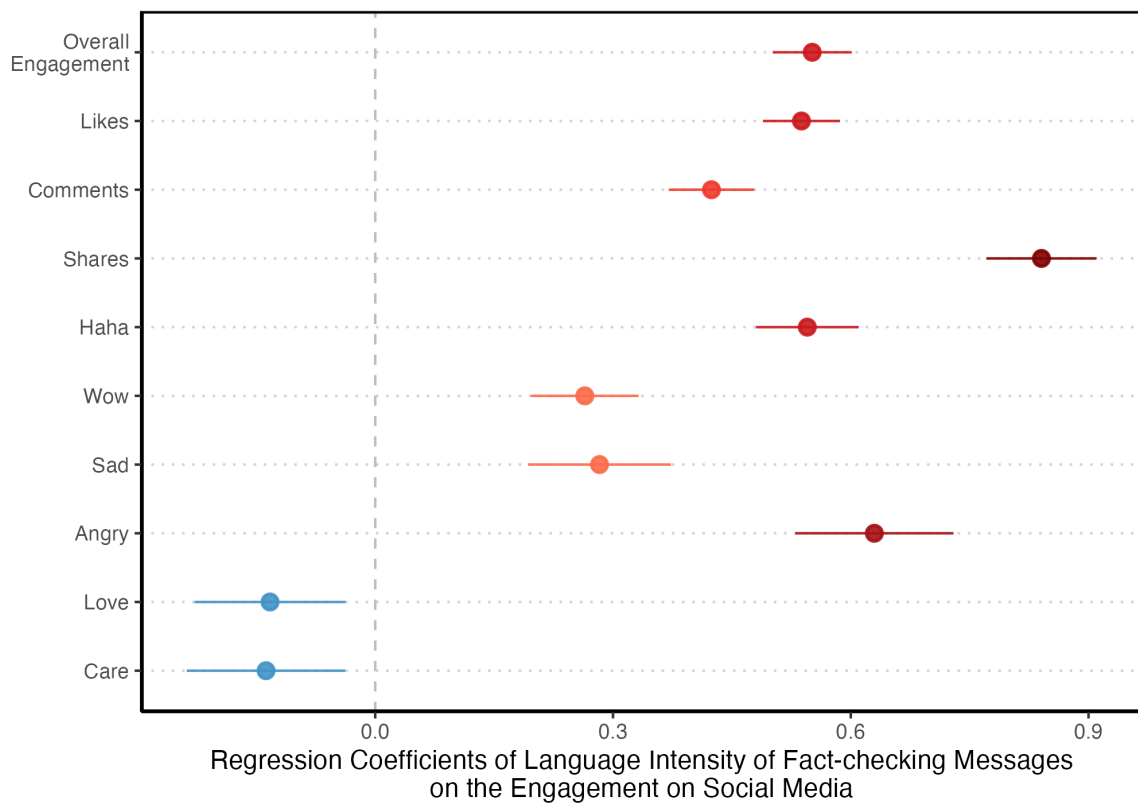Figure 1. Flow diagram of the data collection process.

Figure 2. Negative binomial regression coefficients and confidence intervals of language intensity on the overall engagement and 9 Facebook engagement metrics ($N = 35,487$). The overall engagement is a composite score of all 9 Facebook engagement metrics. In the negative binomial models, language intensity was the independent variable; fact-checking message length, fact-checking agency type (independent, in-house), the number of followers at posting, and the fact-checking message topic were included as covariates.
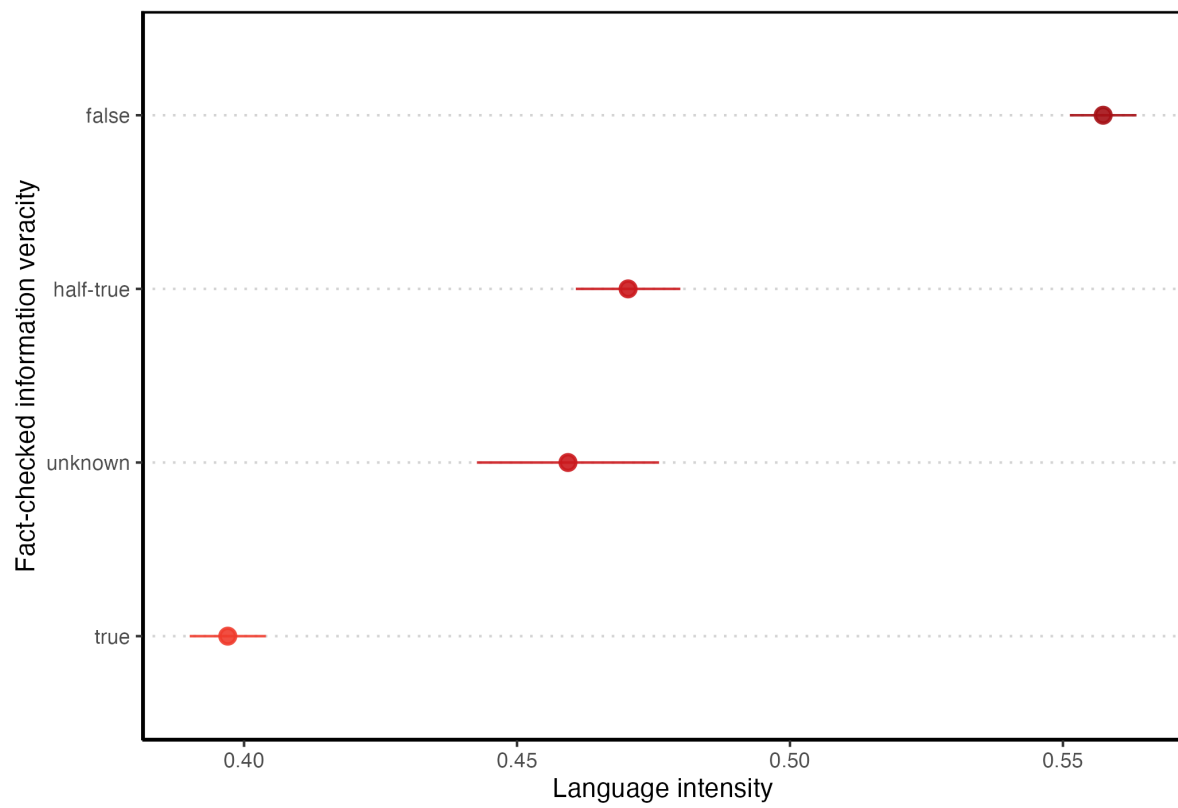
Figure 3. Means and confidence intervals of language intensity across fact-checked information veracity.
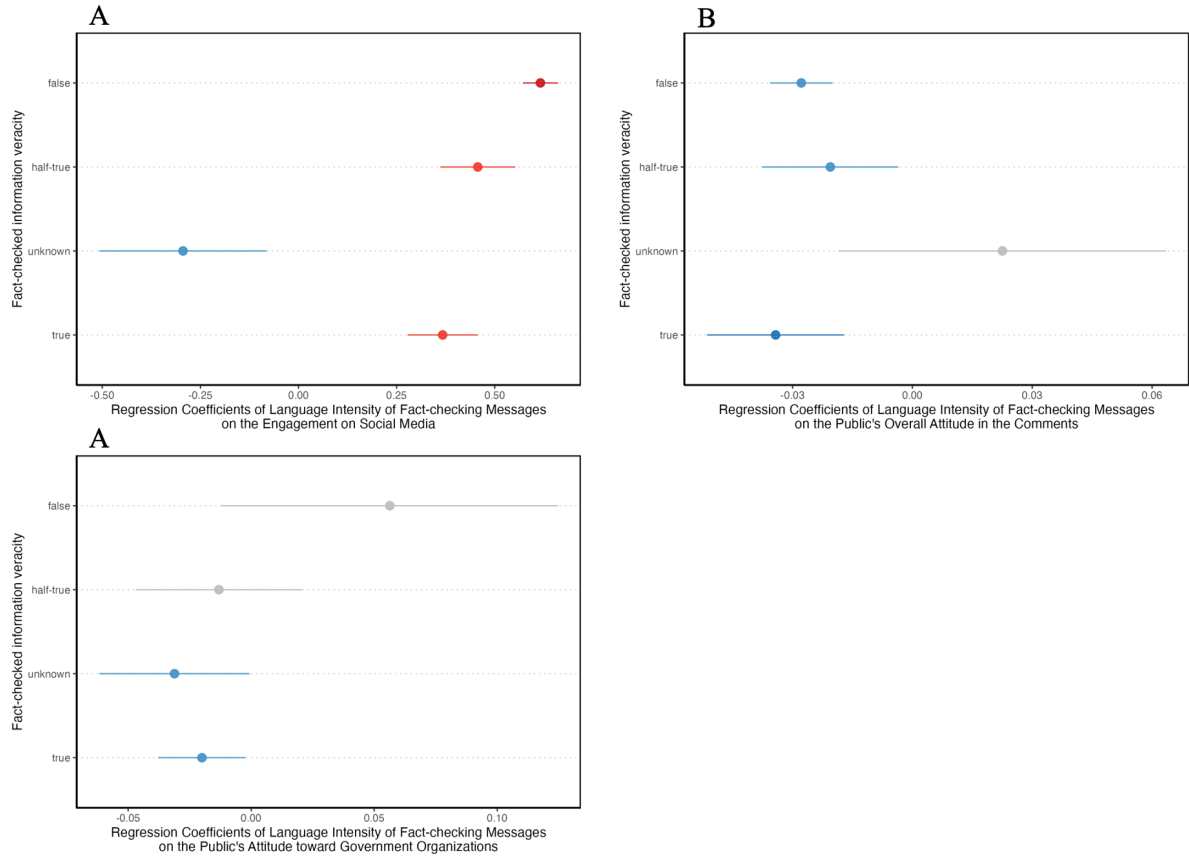
Figure 4. (A) Negative binomial regression coefficients and confidence intervals of language intensity on the overall engagement across fact-checked information veracity. (B) Linear regression coefficients and confidence intervals of language intensity on the public's overall attitude in the comments attached to fact-checking messages. (C) Linear regression coefficients and confidence intervals of language intensity on the public's attitude toward government organizations. In these models, language intensity was the independent variable; fact-checking message length, fact-checking agency type (independent, in-house), and the number of followers at posting were included as covariates.