

Authors: Ronja van Zijverden, Marloes van Moort, Karin Fikkers, Antal van den Bosch, Hans Hoeken

Interactive Paper (Poster) Session

The language of likes: A corpus study on linguistic predictors of text agreement

Although linguistic features are shown to influence text agreement, studies have typically examined only a few features at a time (e.g., Bullock et al., 2019; Nadarevic et al., 2021). Therefore, it remains unclear which combination of features shapes text agreement. To address this, we analyzed a corpus of featured (i.e., pinned) comments (≥ 75 words; $N = 2567$) on articles from a well-known Dutch news website to assess which linguistic features best predict text agreement (operationalized as the number of likes comments received; e.g., Sumner et al., 2018).

Using *T-scan* (Pander Maat et al., 2014), we extracted 447 linguistic features across eight classes: lexical complexity, sentence complexity, referential cohesion and lexical diversity, relational coherence, concreteness, personal style, verbs and time, and parts-of-speech. We created a training and test set (90/10 split) and performed a ten-fold cross-validation with feature selection using correlation filtering, multicollinearity checks, and backward stepwise regression to identify which features best predict likes. Model performance was assessed to evaluate its generalizability and the final model was obtained with the full training and test set.

Preliminary results show consistent selection of features in lexical complexity, referential cohesion and lexical diversity, concreteness, and parts-of-speech (specifically, intensifiers), despite variation across folds. More detailed results will be presented at the conference.

This study extends prior work by identifying which text characteristics most strongly affect text agreement, which contributes to our understanding of cues readers rely on when evaluating text credibility and can aid our understanding of what makes texts persuasive.

References

- Bullock, O. M., Colón Amill, D., Shulman, H. C., & Dixon, G. N. (2019). Jargon as a barrier to effective science communication: Evidence from metacognition. *Public Understanding of Science*, 28(7), 845–853.
<https://doi.org/10.1177/0963662519865687>

- Nadarevic, L., Schnuerch, M., & Stegemann, M. J. (2021). Judging fast and slow: The truth effect does not increase undertime-pressure conditions. *Judgment and Decision Making*, 16(5), 1234–1266. <https://doi.org/10.1017/S193029750000841X>
- Pander Maat, H., Kraf, R., Van den Bosch, A. P. J., Dekker, N., Gompel, M. van, Kleijn, S. de, Sanders, T., & Sloot, K. van der. (2014). *T-Scan: A new tool for analyzing Dutch text*. <https://repository.ubn.ru.nl/bitstream/handle/2066/134833/134833.pdf>
- Sumner, E. M., Ruge-Jones, L., & Alcorn, D. (2018). A functional approach to the Facebook Like button: An exploration of meaning, interpersonal functionality, and potential alternative response buttons. *New Media & Society*, 20(4), 1451–1469. <https://doi.org/10.1177/1461444817697917>