

## Variation in lexical sentiment does not always reflect emotional ambivalence

Emotional meaning is central to how words are processed and used, yet most sentiment resources reduce this complexity to a single average score per word. These scores vary - within surveys, across lexica, in contextual usage, and among semantic neighbors - and that variation is usually treated as noise (cf. Plank 2022). We ask whether such variability is actually informative, indicating emotional ambiguity: emotional polysemy (dispersion of valence across meanings) and ambivalence (co-presence of positive and negative meanings). This matters for both theory and practice. If variation reflects genuine ambiguity, it should be considered in cognitive-variationist research and in applications of tools that depend on sentiment lexica.

We analyze 117 German words using four variability sources: (i) variability across 16 sentiment dictionaries covering diverse designs and populations (as surveyed by Kern et al., 2021), (ii) within a new crowdsourced sentiment survey of 49 L1 speakers, (iii) among distributional semantic neighbors from pretrained embeddings, and (iv) across sentence contexts using a BERT-based German sentiment classifier (Guhr et al., 2020) applied to a balanced contemporary corpus. As a gold standard, we elicit subjectively rated ambivalence via crowdsourcing (binary judgments of whether a word can be both positive and negative). We also consider frequency, number of dictionary senses, and part of speech.

We find that mean valence aligns well across all methods, confirming that central tendencies are robust. In contrast, dispersion-based measures diverge. Crucially, subjectively rated ambivalence correlates with variability in human ratings - both across dictionaries and within our survey - but not with neighborhood- or context-based variability. These relationships persist when controlling for frequency and lexical properties. The findings imply that survey-based variation carries genuine information about emotional ambiguity (cf. Kuhlmann et al., 2017) and support a cognitive-variationist view that inter-speaker and inter-survey differences reflect meaningful socio-cognitive diversity rather than mere noise (Tagliamonte, 2011). By contrast, variability inferred from semantic neighbors or contexts - both inherently sensitive to usage distributions - fails to track perceived ambivalence, suggesting that speakers can readily access positive and negative senses even when they are infrequent. This tempers expectations for inferring ambiguity from distributional propagation and contextual sentiment alone, despite their demonstrated utility for estimating mean valence (e.g., Buechel et al., 2016).

Limitations include the modest sample size dictated by German resources and reliance on single embedding and classifier choices. Overall, the study shows that variability in human sentiment annotations is a valuable signal for emotional ambiguity, while distributional and contextual variance should be used with caution for that purpose.

414 words

## References

- Buechel, S., Hellrich, J., & Hahn, U. (2016). Feelings from the past—Adapting affective lexicons for historical emotion analysis. In *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)* (pp. 54–61).
- Guhr, O., Schumann, A. K., Bahrmann, F., & Böhme, H. J. (2020). Training a broad-coverage German sentiment classification model for dialog systems. In *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 1627–1632).
- Kern, B. M., Baumann, A., Kolb, T. E., Sekanina, K., Hofmann, K., Wissik, T., & Neidhardt, J. (2021). A review and cluster analysis of german polarity resources for sentiment analysis. In *3rd Conference on Language, Data and Knowledge (LDK 2021)* (pp. 37-1). Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

- Kuhlmann, M., Hofmann, M. J., & Jacobs, A. M. (2017). If you don't have valence, ask your neighbor: Evaluation of neutral words as a function of affective semantic associates. *Frontiers in Psychology*, 8, 343.
- Plank, B. (2022). The "problem" of human label variation: On ground truth in data, modeling and evaluation. In *Proceedings of the 2022 conference on empirical methods in natural language processing* (pp. 10671-10682).
- Tagliamonte, S. A. (2011). *Variationist sociolinguistics: Change, observation, interpretation*. Wiley-Blackwell.