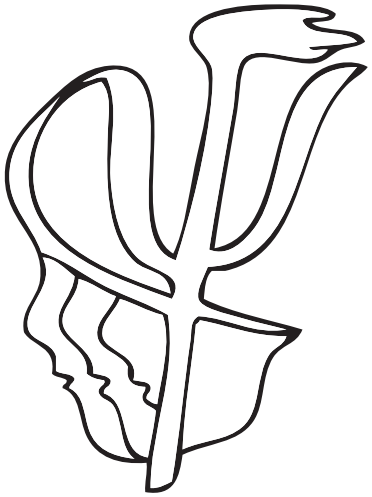

Words, Vectors, and Feelings: Advancing Psychological Emotion Research Through Natural Language Processing

MSc Hubert Wojciech Plisiecki
Graduate School for Social Reseach
Institute of Psychology
Polish Academy of Sciences

Doctoral Thesis prepared under the supervision of
dr hab. Artur Pokropek
and auxillliary supervision of dr. Grzegorz Pochwatko

Warsaw, Poland, 2025

Funding information: The articles published as a part of this thesis were supported by the SONATA BIS grant from the Polish National Science Center "Research Lab for the Digital Social Sciences" (2020/38/E/HS6/00302) led by dr hab. Artur Pokropek

Abstract

To date, psychological research that relies on natural language processing (NLP) has borrowed largely unmodified tools from computer science, often overlooking psychological nuances that are crucial to realizing their full potential for the field. This dissertation aims to correct this pitfall by creating new NLP methods tailored to psychological inquiry, addressing domain-specific concerns such as ecological validity, construct bias, and measurement fidelity, thereby opening new avenues for theory, application, and measurement in affective science.

The first study shows that the fundamental structure of emotion can be recovered directly from everyday language. Using 58 000 Reddit posts tagged with 28 emotions, it builds unsupervised numerical emotion representations and submits them to principal-component analysis. By demonstrating that the circumplex model emerges organically from natural discourse, the study tackles the domain-specific concern of whether laboratory-derived structures generalise to "in-the-wild" expression and provides a blueprint for theorists to mine large corpora for latent psychological dimensions, expanding both theorical and exploratory potential of text-based studies.

The second study extends affective-norm databases with transformer models fine-tuned on existing ratings. The resulting models predict a range of word-level affective indices for unseen words with correlations up to $r = 0.95$ on the test set in English and comparably high scores in five other languages. In doing so, it addresses the practical bottleneck of sparse normative data - a core domain-specific limitation - and introduces automated extrapolation and stimuli-descent procedures that enable researchers to design better-controlled experiments.

The third study audits social bias in a novel Polish sentiment model. Regression analyses reveal that a politician's gender and party affiliation account for up to 66% of the variance in predicted valence—an effect rooted in the annotators' personal biases. By empirically exposing how supervised NLP pipelines can silently propagate psychologically relevant social and political biases, the work opens new methodological avenues for diagnosing and correcting such biases, thereby safeguarding the integrity of psychological measurement and scientific conclusions.

The fourth study introduces the Semantic Propagation Graph Neural Network, an explainable sentiment model designed to retain accuracy while curbing social bias. By "blinding" the network to word identity and letting the emotional information of singular words flow only through syntactic links, the model approaches transformer-level performance and significantly reduces prediction bias, demonstrating that high accuracy and fairness can coexist. This architecture directly addresses the domain-specific concern of

balancing validity with ethical neutrality and offers a potential method of tracking how emotional meaning propagates through syntax, broadening future applications of graph-based modelling in psychology.

Collectively, these four studies contribute a vetted battery of exploratory, predictive, diagnostic, and corrective methods that enable psychologists to investigate mind and behaviour through text while guarding against the methodological and ethical pitfalls of off-the-shelf NLP tools. Together they respond to pressing domain-specific challenges - ecological validity, normative coverage, bias mitigation, and interpretability - while opening fresh trajectories for future psychological theory, experimental application, and large-scale, text-based measurement.

***Keywords*:** Emotion Research, Natural Language Processing, Affective Norm Extrapolation, Social Bias in Sentiment Analysis, Psychological Methods