What Words Reveal About Mental Health: A Computational Language Analysis Around Phase Transitions in Psychotherapy

Mateja Šutar
University of Ljubljana
Ljubljana, Slovenia
mateja.sutar@gmail.com

Tine Kolenik
Institute of Synergetics and Psychotherapy Research
Paracelsus Medical University
Salzburg, Austria
tine.kolenik@ccsys.de

Günter Schiepek
Institute of Synergetics and Psychotherapy Research
Paracelsus Medical University
Salzburg, Austria
guenter.schiepek@ccsys.de

Wolfgang Aichhorn
University Hospital of Psychiatry, Psychotherapy, and Psychosomatics
Salzburg, Austria
w.aichhorn@salk.at

Abstract

Language can reflect key psychological changes during psychotherapy, known as phase transitions (PTs). These sudden shifts in mood, insight, or symptom severity are often expressed in clients' written narratives. We investigated how linguistic features in client diaries relate to PTs by combining textual data with clinical assessments. Feature changes were analyzed using within-participant comparisons and aggregated group-level analysis. Results revealed systematic shifts in word count, pronoun use, and psychological processes-related terms surrounding PTs. These findings may offer additional insight into therapeutic progress and support the development of novel interventions.

Keywords

language use, linguistic shifts, LIWC, phase transitions, psychotherapy, mental health

1 Introduction

Language is first and foremost a tool for communication, enabling humans to share ideas, emotions, and knowledge [1]. In turn, everyday language carries subtle cues about our psychological states, which researchers have long analyzed to gain insight into thought and behavior. Beyond its role in communication, linguistic behavior reflects underlying mechanisms of attention, affect regulation, and self-concept, making it an increasingly valuable marker in psychology [2]. Recent advances in computational linguistics have demonstrated that distinctive linguistic patterns can serve as proxies for a wide range of mental distress [3], and even psychiatric diagnoses [4]. Thus, language is not only a medium for therapeutic exchange but also a temporal reflection of a person's mental change.

A growing body of research conceptualizes psychotherapy as a complex dynamic system in which sudden, discontinuous changes—commonly referred to as phase transitions (PTs)—signal shifts in a client's psychological state. Such transitions may involve sudden alterations in affective tone, the emergence of new insights, or changes in symptom severity [5]. While quantitative time-series approaches, such as the analysis of questionnaires, have shed light on the temporal dynamics of PTs, far less is known about how these key points are manifested in patients' own narratives. Diary writing, in particular, provides a rich, ecologically valid record of subjective experience, yet the systematic study of its content during psychotherapy remains limited.

Our work addresses this gap by applying computational linguistic methods to patient diaries collected during inpatient psychiatric treatment. Specifically, we examine whether linguistic features change systematically around clinically identified PTs. By integrating text analysis with validated psychometric methods, we aim to explore the content of psychological transitions.

2 Methods

2.1 Participants and Dataset

Our research initially included 28 clients undergoing inpatient psychotherapy; however, one case was excluded due to missing data around phase transitions, resulting in a final sample of 27 anonymized clients. The duration of data collection for each client ranged from 74 to 154 consecutive days of hospitalization, with an average length of 88.3 days. The dataset consisted of daily client diary entries alongside Therapy Process Questionnaire (TPQ) results annotated with clinically determined PTs. In total, 102 PTs were identified, corresponding to a mean of 3.5 PTs per client. The number of PTs per participant ranged from 0 to 5, with all but one participant exhibiting at least one PT. All diary entries were written in German language. Participants entered their diary data digitally via PCs, tablets, or smartphones, with no mention of specific instructions regarding length, content, or frequency beyond daily reporting. TPQ represents a validated self-report measure designed to capture fluctuations in therapeutic progress and symptomatology. Clinical experts independently identified PTs by detecting discontinuities in the TPQ time series. These PTs served as reference points around which we examined changes in language use, allowing us to investigate how linguistic patterns correspond to shifts in clients' psychological states.

Information Society 2025, 6–10 October 2025, Ljubljana, Slovenia

http://doi.org/10.70314/is.2025.cogni.7

2.2 Text preprocessing and Feature Extraction

Diary entries were analyzed using the Linguistic Inquiry and Word Count (LIWC) application [6], which classified words into psychologically relevant categories (e.g., emotion, cognitive processes, time orientation). This procedure yielded 117 extracted features per diary entry, representing both linguistic dimensions (e.g., pronoun use, total function words) and psychological processes (e.g., emotion, cognition, drives). To account for interindividual variability in diary length, all features were normalized as relative frequencies.

2.3 Statistical analysis

To examine linguistic change in the context of PTs, we defined temporal windows of 3, 5, and 7 calendar days before and after each clinically identified transition. At present, there is little empirical guidance on how to determine the appropriate time frame for detecting language shifts during psychotherapy. Prior research on linguistic responses to traumatic events, however, suggests that linguistic changes are often immediate but short-lived. For instance, following the 9/11 attacks, the diaries of an on-line journaling service revealed sharp increases in negative emotion, cognitive engagement, and social referencing that largely returned to baseline within about a week [7]. Drawing on this evidence, we adopted multiple window sizes to capture both short-term and extended dynamics surrounding PTs, as visualized in Figure 1.

Two levels of analysis were performed:

Within-participant analysis: For each participant, we compared pre- and post-transition feature distributions using the Wilcoxon Signed-Rank Test, a nonparametric test suitable for paired, non-normally distributed data [8]. Given the exploratory nature of this analysis, we adopted a liberal threshold ($p < 0.15$). Each PT was treated separately rather than averaging across a participant's multiple PTs, allowing us to capture transition-specific dynamics.

Aggregated group-level analysis: To identify consistent patterns across participants, pre- and post-transition feature values were aggregated across participants and tested using the Wilcoxon Rank-Sum Test ($p < 0.05$). This approach allowed us to examine group-level patterns, leveraging the summaries from each PT.

By combining individual- and group-level analyses, we aimed to capture both within-person change processes and shared linguistic dynamics indicative of psychotherapeutic turning points.

3 Results

We found no observable changes in linguistic features within the 3-day window in the within-participant analysis. Conversely, several linguistic features showed consistent changes across both the 5-day and 7-day windows. At 5 days, the most frequent individual shifts involved average sentence length (19 PTs, 15 drops, 4 gains), the total number of pronouns (15 PTs, 8 drops, 7 gains), negative emotion (14 PTs, 10 drops, 4 gains), and drives (14 PTs, 9 drops, 5 gains), while the 7-day window showed most frequent changes in all punctuation (18 PTs, 6 drops, 12 gains), average sentence length (17 PTs, 13 drops, 4 gains), word count (17 PTs, 11 drops, 6 gains), and certainty (17 PTs, 8 drops, 9 gains).

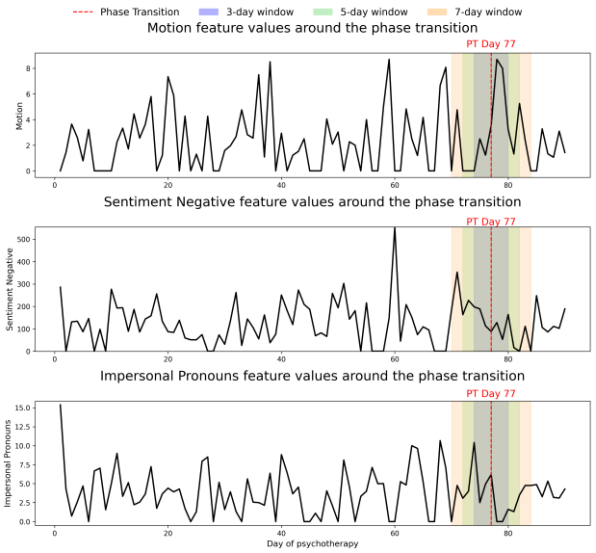

Figure 1: Visualization of linguistic shifts around a client's phase transition (PT). This figure shows shifts in linguistic features (Sentiment Negative, Impersonal Pronouns, Motion) tracked over 90 days of psychotherapy. Red dashed lines mark a PT identified through clinical assessment. Shaded regions represent temporal analysis windows of 3 (violet), 5 (green), and 7 days (orange) before and after each PT. The plots illustrate how different linguistic features may exhibit distinct patterns of change around the same turning point. To illustrate, diary entries corresponding to this specific PT shifted from "Today was a very exhausting day… I notice that I have trouble concentrating…" (PT−1) to "I tried slacklining for the first time… It makes me focus completely and the little successes feel amazing." (PT+5), exemplifying the qualitative change in language accompanying the transition.

Aggregated analysis showed some shared patterns for 5-day and 7-day windows. An overview of the results is presented in Table 1. It includes decreases in achievement- (Δ median -1.64 pp, |r|=0.22, q=0.0863 for 5-day window; Δ median -2.30 pp, |r|=0.37, q=0.000038 for 7-day window), work- (Δ median -1.32 pp, |r|=0.39, q=0.0065 for 5-day window; Δ median -1.44 pp, |r|=0.27, q=0.0077 for 7-day window), feeling-, female-, and power- terms, as well as increases in adverbs (Δ median 1.44 pp, |r|=0.23, q=0.0725 for 5-day window; Δ median 1.50 pp, |r|=0.22, q=0.019 for 7-day window), past focus (Δ median 2.35 pp, |r|=0.35, q=0.0025 for 5-day window; Δ median 3.98 pp, |r|=0.22, q=0.028 for 7-day window), home-terms, and 1st person plural expressions. Unique to the 5-day window were decreases in affect (Δ median -4.19 pp, |r|=0.42, q=0.0021), impersonal pronouns (Δ median -1.96 pp, |r|=0.40, q=0.0021), negative emotion (Δ median -1.18 pp, |r|=0.24, q=0.036), articles, comma use, and reward-terms, while the 7-day window alone showed decreases in drives (Δ median -2.94 pp, |r|=0.20, q=0.023), and discrepancy-terms (Δ median -0.63 pp, |r|=0.14, q=0.12). Increases in differentiation- (Δ median 1.59 pp, |r|=0.22, q=0.073), family- (Δ median 0.40 pp, |r|=0.31, q=0.074), and money-related terms were specific to the 5-day window, while increases in positive emotion (Δ median 5.31 pp, |r|=0.40, q=0.0049), negative emotion (Δ median 1.11 pp, |r|=0.18,

q=0.036), anger (Δ median 0.29 pp, |r|=0.23, q=0.023), personal pronouns (Δ median 3.58 pp, |r|=0.34, q=0.0076), prepositions, conjunctions, negations, netspeak, and time-terms were unique to the 7-day window.

4 Discussion

Our results indicate that measurable language changes occur around phase transitions in clients undergoing psychotherapy. These changes, particularly in content categories, can provide insight into the psychological processes associated with such transitions. Because data were aggregated across diverse participants, the observed patterns were heterogeneous: some participants showed improvement, while others experienced deterioration. This variability likely accounts for the simultaneous increases in both positive and negative emotion features in the aggregated data. Thus, apparent contradictions in directionality may reflect mixed individual trajectories, as the analysis was not grouped by phase transition type.

In our results, several function word categories—such as articles, prepositions, personal pronouns, impersonal pronouns, conjunctions, adverbs, and negations—were also observed. These terms, along with auxiliary verbs, are used in the Analytical Thinking feature, also known as the Categorical-Dynamic Index (CDI) [9], which is a metric of logical thinking. Studies revealed that the CDI reflects students' thinking style and is linked to differences in academic performance [10].

4.1 Language Characteristics of Distinct Mental Health Disorders

Previous studies have documented that different mental health disorders are associated with distinct patterns of language use. For example, ADHD is linked to more third-person plural pronouns and shorter clauses [11, 12], while bipolar disorder shows greater self-focus and references to death [13]. Borderline personality disorder (BPD) involves more swear words, death-related words, and third-person singular pronouns [3]. Individuals with social anxiety disorder (SAD) used self-referential, anxiety, and sensory words, and made fewer references to other people [14]; Major depressive disorder (MDD) involves first-person pronouns, past tense, and repetitive, short sentences [15]. Schizophrenia relates to low semantic cohesion, anger- and religion-related words, references to auditory hallucinations, while also characterized by decreased usage of words related to work, friends, and health [3, 16].

4.2 LIWC Analysis

LIWC is a popular top-down method that offers several advantages for the study of language and cognition. It is a standardized, replicable, and efficient method for quantifying large volumes of textual data to extract psychologically relevant and psychometrically valid measures from language [2, 3]. Top-down methods are based on "dictionaries," categories of words or phrases, each associated with a given construct or set of constructs, such as anxiety or suicidal ideation [2]. This enables researchers to detect subtle emotional and cognitive dynamics that may not be captured with traditional self-report measures, making it a powerful complement to other assessment tools.

Table 1: Aggregated analysis results

Category	Most frequently used examples	Direction (Gain ↑ / Drop ↓)	Time-window
Work	work, school, working, class	↓	5 & 7 days
Achievement	work, better, best, working		
Feeling	feel, hard, cool, felt		
Power	own, order, allow, power		
Female	she, her, girl, woman		
Adverbs	so, just, about, there	↑	
Home	home, house, room, bed		
Past focus	was, had, were, been		
1st person plural	we, our, us, lets		
Negative emotion	hate, bad, hurt, tired	↓	5 days
Impersonal pronouns	that, it, this, what		
Affect	emotion, mood		
Articles	a, an, the, alot		
Reward	opportun*, win, gain*, benefit*		
Comma			
Differentiation	but, not, if, or	↑	
Family	parent*, mother*, father*, baby		
Money	business*, pay*, price*, market*		
Discrep	would, can, want, could	↓	7 days
Drives	we, our, work, us		
Negative emotion	hate, bad, hurt, tired	↑	
Positive emotion	good, love, happy, hope		
Anger	hate, mad, angry, frustr*		
Time	when, now, then, day		
Personal pronouns	I, you, my, me		
Negations	not, no, never, nothing		
Prepositions	to, of, in, for		
Conjunctions	and, but, so, as		
Netspeak	:), u, lol, haha*		

4.2.1 Top-down vs. Bottom-up Methods. Top-down methods, while highly structured, may sometimes overlook context-specific, cultural, or metaphorical nuances [2]. Bottom-up approaches, by contrast, focus on broader patterns in language rather than predefined constructs. Techniques such as probabilistic topic models [17], statistical semantic models [18], and neural language models [19] capture characteristics ranging

from word co-occurrence and meaning to sequential dependencies.

Combining top-down, bottom-up, and qualitative approaches enables a highly nuanced and insightful analysis of textual data. This integrated strategy allows researchers not only to quantify specific psychological constructs but also to examine emergent patterns, contextual nuances, and complex semantic structures, providing a comprehensive understanding of language use and its psychological implications [2].

4.3 Limitations

Interpretation of our findings is limited by the absence of information about clients' diagnoses and annotations regarding the nature of phase transitions, indicating whether the transition represents improvement or worsening of symptoms. Other limitations include heterogeneity of participants, contextual limitations of LIWC, and the absence of fine-grained temporal resolution.

5 Conclusion

Our research suggests that language shifts hold potential as indicators of psychological change. Understanding these patterns may provide clinicians with more sensitive indicators of therapeutic progress, offering potential guidance for interventions, and improving the precision of treatment monitoring in inpatient psychiatric care.

6 Future Work

Future research could implement transformer-based neural network architectures (e.g., BERT, RoBERTa) to cluster participants according to symptom trajectories, such as improvement or deterioration. Analyses could then be conducted to examine differences in linguistic shifts across clusters. Where available, results from neural language models could be compared with clinical annotations to evaluate prediction accuracy. Future studies should aim to link these linguistic patterns more directly to specific mental states, ultimately supporting the development of clinically relevant interventions and applications.

Acknowledgments

This research was supported by Paracelsus Medical University, which also provided access to the clinical dataset utilized in this study. The language of this paper was revised with the assistance of ChatGPT-5.

References

[1] Evelina Fedorenko, Steven T. Piantadosi, and Edward A. F. Gibson. Language is primarily a tool for communication rather than thought. *Nature* 630, 8017 (Jul. 2024), 575–586. DOI: https://doi.org/10.1038/s41586-024-07522-w

[2] [2] Brendan Kennedy, Ashwini Ashokkumar, Ryan L. Boyd, and Morteza Dehghani. 2022. Text analysis for psychology: Methods, principles, and practices. In Handbook of Language Analysis in Psychology. Morteza Dehghani and Ryan L. Boyd (Eds.). The Guilford Press, New York, NY.

[3] Minna Lyons, Nazli D. Aksayli, and Gayle Brewer. Mental distress and language use: Linguistic analysis of discussion forum posts. Comput. Hum. Behav. 87 (Oct. 2018), 207–211. DOI:https://doi.org/10.1016/j.chb.2018.05.035

[4] Marco Spruit, Stephanie Verkleij, Kees de Schepper, and Floortje Scheepers. Exploring language markers of mental health in psychiatric stories. Appl. Sci. 12, 4 (Feb. 2022), 1–17. DOI:https://doi.org/10.3390/app12042179

[5] Günter K. Schiepek, Kathrin Viol, Wolfgang Aichhorn, Marc-Thorsten Hütt, Katharina Sungler, David Pincus, and Helmut J. Schöller. Psychotherapy is chaotic—(not only) in a computational world. Front. Psychol. 8 (Apr. 2017), 379. DOI:https://doi.org/10.3389/fpsyg.2017.00379

[6] Ryan L. Boyd, Ashwini Ashokkumar, Sarah Seraj, and James W. Pennebaker. The development and psychometric properties of LIWC-22. University of Texas at Austin, Austin, TX. https://www.liwc.app

[7] Michael A. Cohn, Matthias R. Mehl, and James W. Pennebaker. Linguistic markers of psychological change surrounding September 11, 2001. Psychol. Sci. 15, 10 (Oct. 2004), 687–693. DOI:https://doi.org/10.1111/j.0956-7976.2004.00741.x

[8] Bernard Rosner, Robert J. Glynn, and Mei-Ling T. Lee. The Wilcoxon signed rank test for paired comparisons of clustered data. Biometrics 62, 1 (Mar. 2006), 185–192. DOI:https://doi.org/10.1111/j.1541-0420.2005.00389.x

[9] Boban Simonovic, Katia Correa Vione, Edward Stupple, and Alice Doherty. It is not what you think it is how you think: A critical thinking intervention enhances argumentation, analytic thinking and metacognitive sensitivity. Think. Skills Creat. 49 (Jun. 2023), 101362. DOI:https://doi.org/10.1016/j.tsc.2023.101362

[10] James W. Pennebaker, Cindy K. Chung, Joey Frazee, Gary M. Lavergne, and David I. Beaver. When small words foretell academic success: The case of college admissions essays. PLoS One 9, 12 (Dec. 2014), e115844. DOI:https://doi.org/10.1371/journal.pone.0115844

[11] Glen Coppersmith, Mark Dredze, Craig Harman, and Kristy Hollingshead. From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. 2015. In Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, CLPsych 2015. June 5, 2015, Denver. Association for Computational Linguistics, Denver, CO, 1–10. DOI:https://doi.org/10.3115/v1/W15-1201

[12] Kyungil Kim, Seongjik Lee, and Changhwan Lee. College students with ADHD traits and their language styles. J. Atten. Disord. 19, 8 (Aug. 2015), 687–693. DOI:https://doi.org/10.1177/1087054712452343

[13] Marie Forgeard. Linguistic styles of eminent writers suffering from unipolar and bipolar mood disorder. Creat. Res. J. 20, 1 (Feb. 2008), 81–92. DOI:https://doi.org/10.1080/10400410701842094

[14] Barrett Anderson, Philippe R. Goldin, Keiko Kurita, and James J. Gross. Self-representation in social anxiety disorder: Linguistic analysis of autobiographical narratives. Behav. Res. Ther. 46, 10 (Oct. 2008), 1119–1125. DOI:https://doi.org/10.1016/j.brat.2008.07.001

[15] Raluca N. Trifu, Bogdan Nemeş, Carolina Bodea-Hategan, and Doina Cozman. Linguistic indicators of language in major depressive disorder (MDD): An evidence-based research. J. Evid.-Based Psychother. 17, 1 (Mar. 2017), 105–128.

[16] Michael L. Birnbaum, Sindhu K. Ernala, A. F. Rizvi, Elizabeth Arenare, Anna Van Meter, M. De Choudhury, and J. M. Kane. Detecting relapse in youth with psychotic disorders utilizing patient-generated and patient-contributed digital data from Facebook. NPJ Schizophr. 5, 1 (Dec. 2019), 17. DOI:https://doi.org/10.1038/s41537-019-0085-9

[17] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent Dirichlet allocation. J. Mach. Learn. Res. 3 (Mar. 2003), 993–1022.

[18] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In Proceedings of the 27th International Conference on Neural Information Processing Systems – Volume 2 (NIPS'13), December 5 - 10, 2013, Lake Tahoe Nevada. Curran Associates Inc., Red Hook, NY, USA, 3111–3119.

[19] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Jauvin. A neural probabilistic language model. J. Mach. Learn. Res. 3 (Mar. 2003), 1137–1155.