A cognitively informed biomarker and assessment tool of positive thought disorder: The strategic use of Large Language Models

Lin Wang, Anthony Yacovone, Tori Sharpe, Sabrina Ford, Lena Palaniyappan, Gina Kuperberg

Background

Positive thought disorder (PTD) is prominent in schizophrenia and characterized by patients' *over*-reliance on local associations and *under*-reliance on the global discourse during language processing [1,2,3,4]. Traditional clinical assessments of PTD (e.g. diagnostics of loosening of associations, derailment, and tangentiality) are subjective and time-consuming. We argue that Large Language Models (LLMs) can be used as objective, automatic assessment tools for measuring PTD. In this study, we used two LLMs, BERTopic [5] and Word2Vec [6], to characterize the nature of global and local representations present in naturally produced speech. We find that these measures successfully capture the linguistic consequences of PTD.

Methods

We analyzed 3 picture descriptions (~1 minute each) from 70 first-episode psychosis patients (all unmedicated and treatment-naïve) and 34 demographically matched controls. We used the Disorganization subscore from the Thought and Language Index (TLI) [7] to measure PTD. Symptoms and domain-general cognitive function were assessed using the PANSS-8 [8], Semantic Fluency, Digit-Symbol Substitution, and the Trail-Making Test.

Results

***Model 1* (BERTopic):** We used a pre-trained BERTopic model with 2376 common topics ascertained from the Wikipedia corpus [9] to assess how "topics" shifted throughout each picture description. Specifically, for each participant, we ascertained the variability in their topic representation, and characterized that variability as entropy. We then regressed these values onto TLI (while controlling for description length, age, gender, and SES). We found that TLI significantly predicted topic entropy such that participants with greater TLI scores produced more topics in their descriptions ($b = 0.06$, $SE = 0.03$*,* $t = 2.17$, $p = 0.03$).

***Model 2* (Word2Vec):** We used a pre-trained Word2Vec model with 300-dimensional word embeddings [6] to calculate the pairwise cosine similarity for a given content word with each of its five preceding content words. For our analyses, we averaged similarity values across all content words for each pairing. We then regressed similarity values onto TLI, word pairing, their interaction, and the control variables above. We found that larger TLI scores resulted in greater similarities between words and their local contexts ($b = 0.01$*,* $SE = 0.003$, $t = 2.88$, $p = 0.005$), and this similarity begins to weaken at 4 words back.

***Model 3* (Classification):** We also asked if entropy and cosine similarity values can jointly discriminate between patients and controls. To do this, we divided the data into training and testing sets (80-20), and trained a model to predict patients and controls using their entropy and

similarity values. The model can effectively distinguish between patients and controls: overall accuracy = 0.86, sensitivity = 1.0, precision = 0.83, and AUC = 0.91.

Discussion

This work has clinical, neurocognitive, and computational implications: LLMs can provide linguistic biomarkers for fast, automated, and objective quantification of language disorganization, as PTD predicts the entropy and local similarity in patients' speech. LLM biomarkers can be used for illness detection, symptom monitoring, and outcome predictions. These results bridge the clinical characterizations of PTD with neurocognitive evidence for selective atypicalities in patients' language processing. Our work supports hierarchical generative models of psychosis, and future work will situate these findings within the biologically plausible, Predictive Coding framework, to provide a mechanistic account for *over*-reliance on local associations and *under*-reliance on the global discourse during language processing.

References

1. Bleuler, E. *Dementia Praecox, or the Group of Schizophrenias*. (International Universities Press, 1911/1950).
2. Chaika, E. A linguist looks at 'schizophrenic' language. *Brain Lang.* **1**, 257-276 (1974).
3. Andreasen, N. C. Thought, language and communication disorders. II. Diagnostic significance. *Archives of General Psychiatry* **36**, 1325-1330 (1979).
4. Andreasen, N. C. Thought, language and communication disorders. I. Clinical assessment, definition of terms, and evaluation of their reliability. *Archives of General Psychiatry* **36**, 1315-1321 (1979).
5. Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*(arXiv:2203.05794). arXiv. https://doi.org/10.48550/arXiv.2203.05794
6. Mikolov, T., Chen, K., Corrado, G. S., & Dean, J. (2013). Efficient estimation of word representations in vector space. 1st International Conference on Learning Representations (ICLR), Workshop Track Proceedings, Scottsdale, Arizona.
7. Liddle, P. F. *et al.* Thought and Language Index: an instrument for assessing thought and language in schizophrenia. *British Journal of Psychiatry* **181**, 326-330, doi:10.1192/bjp.181.4.326 (2002).
8. Lin, C. H., Lin, H. S., Lin, S. C., Kuo, C. C., Wang, F. C., & Huang, Y. H. (2018). Early improvement in PANSS-30, PANSS-8, and PANSS-6 scores predicts ultimate response and remission during acute treatment of schizophrenia. *Acta Psychiatrica Scandinavica*, *137*(2), 98-108.
9. https://huggingface.co/MaartenGr/BERTopic_Wikipedia