

Introduction

- Positive thought disorder (PTD):
 - Common disabling symptom of schizophrenia
 - Characterized by disorganized language output
- Characterization of PTD during language processing:
 - Over-reliance on associations between neighboring words
 - Under-reliance on information from the broader context
 - Imbalance between *local* and *global* contextual cues
- Traditional clinical assessment tools [1]
 - Loosening of associations, derailment, and tangentiality
 - Subjective and time-consuming
- Natural Language Processing (NLP) tools
 - Objective and automatic
 - But difficult to interpret and largely used for classification
- Present study uses these NLP tools as *theoretical* tools, make them more interpretable and clinically applicable.

Methods

- Participants**
 - 70 individuals with first-episode psychosis (FEP)
 - 34 healthy controls (HC)
- Picture description task**
 - Three Thematic Apperception Test images (~1 min each)
 - Responses were recorded and transcribed
- Clinical and cognitive assessments**
 - Thought and Language Index (TLI) [2]
 - Disorganization metric** = positive thought disorder
 - Impoverishment metric** = negative thought disorder
 - Positive and Negative Syndrome Scale-8 Item (PANSS-8)
 - Semantic Fluency** = Semantic memory
 - Digit-Symbol Substitution** = Working memory
 - Trail-Making Test (Part B)** = Executive function
- Statistical analyses**
 - Linear regressions on two NLP measures: *topic entropy* (global) and *Word2vec similarity* (local)
 - Controls for *gender*, *age*, *SES*, and *description length*
 - Classification of FEP and HC using logistic regressions

Word association: Word2vec

Fig. 1. Compute Word2vec similarity between every word *n* and its five preceding content words.

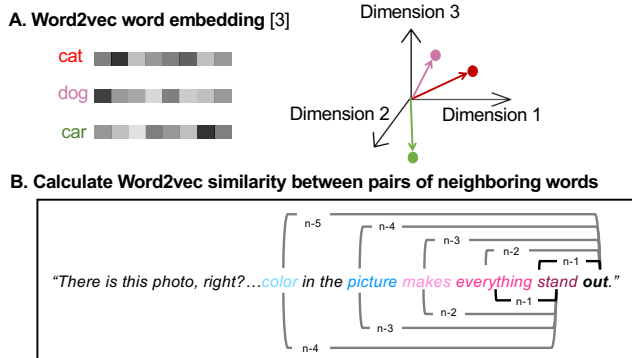
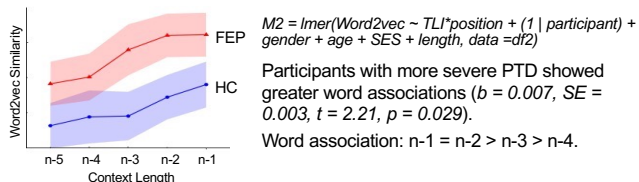
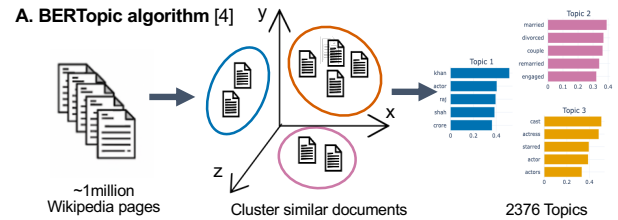


Fig 2. Word2vec similarity between word *n* and its five preceding words.



Topic representation: BERTopic

Fig. 3. Compute topic entropy using BERTopic.



B. Quantify topic distribution: (1) Divide each text into overlapping segments of ~30 words; (2) Quantify similarity between each segment (S) and each of the 2376 topics; (3) Sum and normalize similarity across all segments; (4) Compute topic entropy.

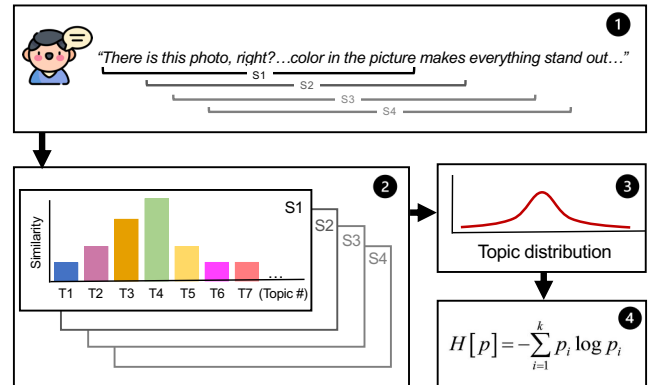
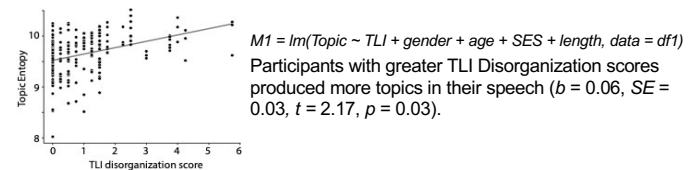
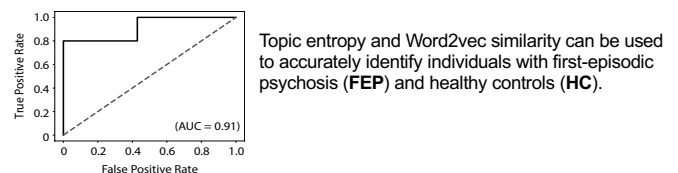


Fig 4. The TLI disorganization score predicts topic diversity in speech.



Classification of FEP vs. HC

Fig 5. ROC curve for classification of FEP vs. HC.



Discussion

- Clinical implications in using Word2vec and BERTopic to quantify language disorganization:
 - Automatic and objective linguistic biomarker
 - Subtypes of PTD:** loosening of association, derailment, and tangentiality
- Bridging clinical observations with neurocognitive evidence for selective deficits in processing global vs. local information in language comprehension.
- Towards a hierarchical generative model of psychosis [5]
 - Uncertainty in the representation of high-level information might be compensated by over-reliance on local associative low-level information.

References

- [1] Andreasen, N. C. Archives of General Psychiatry 36, 1315-1330 (1979); [2] Liddle, P. F. et al. British Journal of Psychiatry 181, 326-330 (2002); [3] Mikolov, T. et al. ICLR Workshop (2013); [4] Grootendorst, M. arXiv:2203.05794 (2022); [5] Brown, M. & Kuperberg, G. Frontiers in Human Neuroscience 9, 643 (2015)