Altered Semantic Similarity in Schizotypy and Psychosis: Evidence from Predictive Language Processing and LLM-Based Analysis

Franziska Knolle

Department of Diagnostic and Interventional Neuroradiology, School of Medicine and Health, Technical University of Munich, Munich, Germany
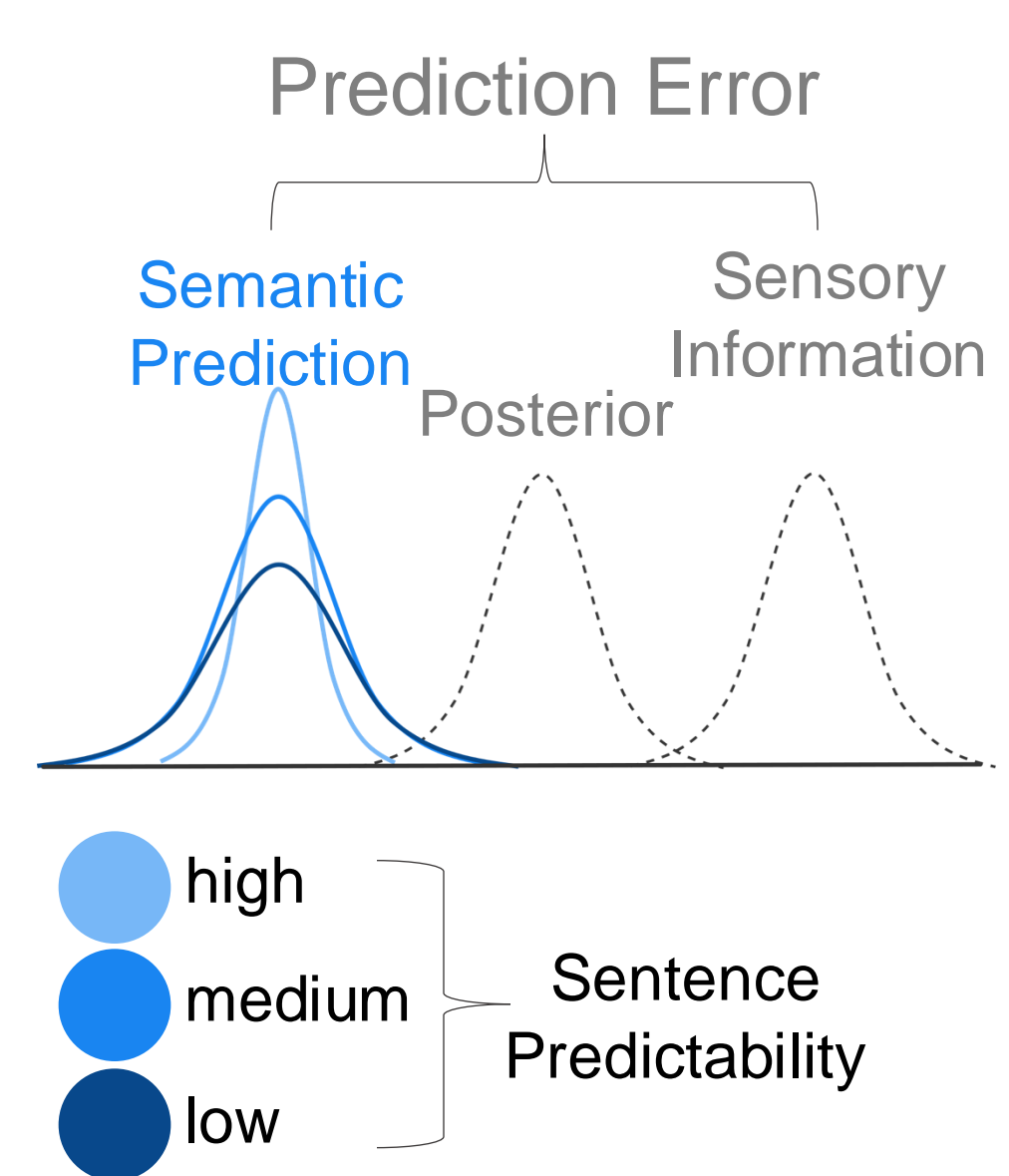
Background

- **Language disturbances** are a core feature of schizophrenia: **semantic processing deficits** – overt **disorganization.**[1]
- Successful language processing, especially during environmental uncertainty, relies on the generation of **semantic predictions**.[2]
- **Predictive Coding** theories report an **overreliance on higher-level semantic predictions** across different disease stages.[3]

Question

Do individuals with increased schizotypal traits and patients with acute psychosis compared psychotic remission generate semantic predictions from a larger semantic space during sentence processing?

Methods

Predictive Coding Theory

- high
- medium
- low

Sentence Predictability

Samples

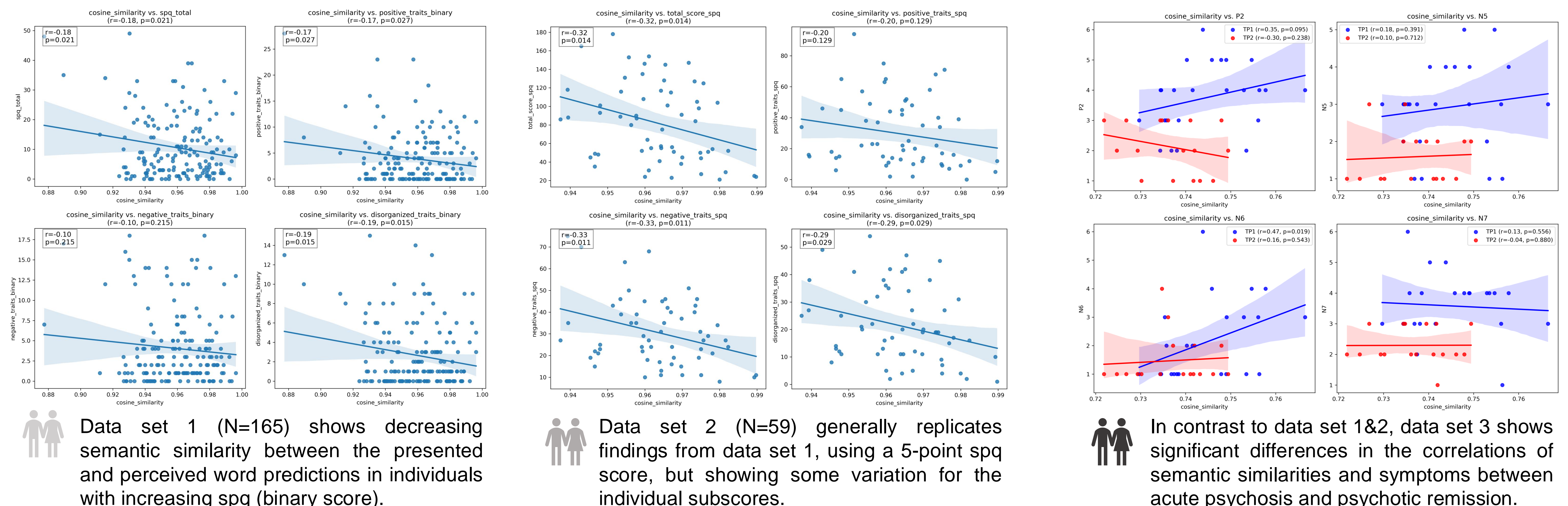|	
N = 165 108f, 57m	**SPQ (0-49)** Mean = 10.64 SD = 10.56
N = 59 37f, 20m	**SPQ (21-178)** Mean = 81.80 SD = 42.38
SCZ = 17 5f, 12m **CON = 20** 8f, 12m	**PANSS pos** Acute psychosis: Mean = 22.0 SD = 2.68, Psy. Remission: Mean = 12.0 SD = 2.61

LLM Approach

- Pre-trained **German BERT model** and tokenizer
- Extract **high dimensional semantic information** (embeddings) for presented and perceived words
- **Cosine Similarity** between the embeddings of presented and perceived words (cosine of the angle between two vectors in a multi-dimensional space)

Predictive Language Paradigm

Condition	Semantic Prior *Entropy*	ISI 100 ms	Sensory Evidence *Cloze* *Sensory Degradation*					ISI 100 ms	Clarity Rating	Posterior Recording	Confidence Rating
Low Entropy (N = 40)	The ship disappeared into the thick *(0)*	+	fog *(1)*	12	6	3	1	+	1 - 100		1 - 100
Medium Entropy (N = 80)	The airplane crashed into a *(1.29)*	+	mountain *(0.53)*	1	12	6	3	+	1 - 100		1 - 100
High Entropy (N = 40)	The new laptop was on top of a *(2.38)*	+	box *(0.05)*	3	1	12	6	+	1 - 100		1 - 100
Low Entropy MM (N = 40)	He drank his coffee with milk and *(0.41)*	+	wine *(0)*	6	3	1	12	+	1 - 100		1 - 100

Semantic Similarity Analysis: Preliminary Findings

Data set 1 (N=165) shows decreasing semantic similarity between the presented and perceived word predictions in individuals with increasing spq (binary score).

Data set 2 (N=59) generally replicates findings from data set 1, using a 5-point spq score, but showing some variation for the individual subscores.

In contrast to data set 1&2, data set 3 shows significant differences in the correlations of semantic similarities and symptoms between acute psychosis and psychotic remission.

Discussion

Lower semantic similarity in subclinical schizotypy may reflect a broader semantic space (wider/global priors), contrasting with acute psychosis (narrower/local priors). These findings suggest a shift from earliest to chronic disease stages possibly. Furthermore, later stages may also be impacted by a stronger cognitive decline.

Next steps

- Updated language model (GPT4) to improve characterization of semantic information extraction.
- Analysis by entropy condition to investigate the impact of sentence predictability.
- Mixed-model analysis to combine integrate all tasks aspects (e.g. channel number) and random effects.

References

1. Kuperberg GR. Language in Schizophrenia Part 1: An Introduction. *Lang Linguist Compass.* 2010;4(8):576-589.
2. Kuperberg GR, Jaeger TF. What do we mean by prediction in language comprehension? *Lang Cogn Neurosci.* 2016;31(1):32-59.
3. Knolle F, Sterner E, Demler VF, MacGregor LJ, Mathys C. Guided by Expectations: Overweighted Semantic Priors in Schizotypy and their Links to Glutamate. *PsyArXiv*, 2024.

franziska.knolle@tum.de
@franziskaknolle.bsky.social