



OPEN ACCESS

Edited and REVIEWED BY
Stefan Borgwardt,
University of Lübeck, Germany

*CORRESPONDENCE

Olivier Dufor

✉ olivier.dufor@isen-ouest.yncrea.fr

Christophe Lemey

✉ christophe.lemey@chu-brest.fr

RECEIVED 15 July 2025

ACCEPTED 28 July 2025

PUBLISHED 13 August 2025

CITATION

Dufor O, Nikzad AH, Lucarini V and Lemey C (2025) Editorial: Natural language processing and artificial intelligence tools to explore the relationship between language and schizophrenia from diagnosis to care. *Front. Psychiatry* 16:1666275. doi: 10.3389/fpsy.2025.1666275

COPYRIGHT

© 2025 Dufor, Nikzad, Lucarini and Lemey. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Natural language processing and artificial intelligence tools to explore the relationship between language and schizophrenia from diagnosis to care

Olivier Dufor^{1*}, Amir Hossein Nikzad^{2,3}, Valeria Lucarini^{4,5,6} and Christophe Lemey^{7,8,9*}

¹L@bISEN, ISEN Yncréa Oeste, Caen, France, ²Institute of Behavioral Science, Feinstein Institutes for Medical Research, Northwell Health, Manhasset, NY, United States, ³Zucker Hillside Hospital, New York, NY, United States, ⁴Institute of Psychiatry and Neuroscience of Paris (IPNP), INSERM U1266, Team: Pathophysiology of Psychiatric Disorders: Development and Vulnerability, Université Paris Cité, Paris, France, ⁵GHU Paris Psychiatrie et Neurosciences, CJAAD, Evaluation, Prevention and Therapeutic Innovation Department, Hôpital Sainte Anne, Paris, France, ⁶CNRS GDR 3557-Institut de Psychiatrie, Paris, France, ⁷URCI University Hospital, Department of Adult Psychiatry, Brest, France, ⁸Sorbonne University, INSERM, Pierre-Louis Institute of Epidemiology and Public Health, Paris, France, ⁹IMT Atlantique, Lab-STICC, Technopôle Brest-Iroise, Brest, France

KEYWORDS

schizophrenia, natural language processing, linguistic markers, psychosis, early detection

Editorial on the Research Topic

Natural language processing and artificial intelligence tools to explore the relationship between language and schizophrenia from diagnosis to care

Schizophrenia is a heterogeneous disorder classically defined by three symptom clusters: positive symptoms (such as hallucinations and delusions), negative symptoms (including affective flattening, avolition, and social withdrawal), and disorganization symptoms (notably thought disorder and incoherent speech). These manifest notably in language disturbances characterized by fragmented, disorganized, or impoverished speech, alongside motor and praxis difficulties as well as impaired social interactions, reflecting underlying neural circuit disruptions (1). Several studies have shown that speech behavior in particular, plays a central role in the symptomatology and diagnosis of patient outcomes (2, 3). A large body of clinical evidence suggests that early detection of schizophrenia spectrum disorder (SSD) and adequate care management and psychosocial interventions can help patients recover from the disease and integrate with the community (4, 5).

In this context, the evaluation of spoken language of patients with SSD or at risk of developing schizophrenia has repeatedly demonstrated its prognostic and diagnostic value (6, 7).

As a proxy for mental activity, language disorders, which represent most of the expression of “formal thought disorders”, manifest themselves through disorganization of speech, loss of coherence, and alteration of emotional expression (8). With natural language processing techniques (NLP), a whole collection of new features appears to contribute to the clinical picture of SSD and clinical high risk for psychosis (CHR-P). While semantic coherence emerge as the principal marker in many studies (9), others emphasized on emotional prosody (10). Nevertheless, most linguistic markers, when considered independently, lack of specificity for schizophrenia and symptomatic alterations go beyond the simple diagnostic framework of schizophrenia or psychosis (11). Then, a multimodal approach, integrating linguistic data with other clinical and biological markers, holds promises to enhance the accuracy and richness of detection and assessments of at risk patients (12, 13).

The objective of this topic was to bring together the most advanced studies in the field of diagnosing and predicting the future of patients with SSD and/or CHR-P from linguistic markers and to think about how they can be mixed in different pathological contexts and trajectories.

Among the articles that make up this topic, that of Just et al., highlights semantic incoherence, meaningly the inability to maintain a logical thread in the discourse. The authors demonstrate that semantic incoherence is one of the key elements in non-affective psychosis that could help in diagnosis. On top of negative symptoms scores which correlate with coherence independently of the embedding model, inpatient care, disorganized score and excitement score add up to the board when Word2Vec method is used.

In addition to the semantic inconsistency, whose predictive value will be discussed later, the article by Olson et al., demonstrates that the tone of the discourse can also be important.

By using Linguistic Inquiry and Word Count, they determine that despite any significant differences in the count of emotionally charged terms, the tone of speech becomes more “negative” in CHR-P patients, particularly if their positive symptoms are high. This result perfectly represents the subtlety of language and the complex interactions between its various levels. Both in the clinical interview and classification criteria, the assessment of emotional states is an integral part of the diagnostic process. In clinical practice, these are spontaneously perceived and often identified through vocal expression, particularly in prosody, intonation, rhythm, or even intensity. To build on these results, it is interesting to remember that paralinguistic abnormalities linked to emotion (e.g., monotone voice, prosody flattening) (14–16) probably reflect affects in a complementary way to verbal content when detection of psychiatric disorders is at stake.

According predictive value, the study by Kim-Dufor et al., uses transcripts of free speech interviews as input to a machine learning model (XGBoost) (17) to automatically classify with 82% accuracy success patients into three categories: not at risk, at risk, and first psychotic episode. The authors also examine the respective contribution to the classification of linguistic markers in the transcribed speech and conclude that semantic coherence,

frequency of pronoun “I,” and filled pauses help in predicting patient’s outcome. This approach reconciles algorithmic performance and clinical intelligibility, providing a more transparent “black box,” which constitutes an essential condition for the acceptability of AI tools in everyday psychiatric practice.

Finally, the relationships between the aforementioned linguistic anomalies and their neural correlates have also been studied in this topic. Applying the PRISMA method on 37 imaging studies, Alonso-Sánchez et al., explored the link between linguistic disturbances such as semantic coherence, maximal semantic coherence or disorganization of thoughts and brain alterations. Thus, whether patients are at ultra-high risk (UHR), have already had a first episode of psychosis (FEP) or present SSD, structural and functional modifications appear. These are mainly driven by differences in processing both in production or comprehension of speech when semantics is involved whether analyzed via NLP models or introduced employing specific experimental paradigms. Functional changes are also found related to disorders of encoding and/or word selection; two functions closely intertwined in the construction of a semantically coherent discourse. The pattern of visible changes also seems to extend from patients with FEP to those with schizophrenia, through UHR patients.

Conclusion

The linguistic and cognitive disruptions highlighted throughout this Research Topic not only deepen our understanding of schizophrenia spectrum disorders but also pave the way toward the development of more faithful cognitive models of language processing, potentially surpassing existing connectionist frameworks (18).

Beyond the diagnostic domain, these AI-driven approaches hold great promise by automating language analysis to provide more objective, faster, and cost-effective evaluations than traditional clinical methods. Such tools could contribute to a psychiatry that is more precise, personalized, and predictive, capable of detecting subtle changes well before the onset of severe symptoms. This progress could also, one day, enable innovative applications like the emergence of a digital twin of the brain’s language functions, offering unprecedented insights into psychosis.

It is important to emphasize that artificial intelligence will never replace the clinician but rather become his most valuable ally, especially in the complex and inherently subjective field of mental health. Once known as a “language disease,” schizophrenia today finds in digital language processing an innovative tool for understanding and care, opening new horizons for both research and clinical practice.

Author contributions

OD: Writing – original draft, Writing – review & editing. AN: Writing – original draft, Writing – review & editing. VL: Writing – review & editing. CL: Writing – original draft, Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial

intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. American Psychiatry Association. *Diagnostic and Statistical Manual of Mental Disorders* | *Psychiatry Online* (2025). doi: 10.1176/appi.books.9780890425596.
2. Andreasen NC, Grove WM. Thought, language, and communication in schizophrenia: diagnosis and prognosis. *Schizophr Bull.* (1986) 12:348–59. doi: 10.1093/schbul/12.3.348
3. de Boer JN, van Hoogdalem M, Mandl RCW, Brummelman J, Voppel AE, Begemann MJH, et al. Language in schizophrenia: relation with diagnosis, symptomatology and white matter tracts. *NPJ Schizophr.* (2020) 6:10. doi: 10.1038/s41537-020-0099-3
4. McGlashan TH, Johannessen JO. Early detection and intervention with schizophrenia: rationale. *Schizophr Bull.* (1996) 22:201–22. doi: 10.1093/schbul/22.2.201
5. Riecher-Rössler A, Gschwandtner U, Borgwardt S, Aston J, Pflüger M, Rössler W. Early detection and treatment of schizophrenia: how early? *Acta Psychiatr Scand.* (2006) 113:73–80. doi: 10.1111/j.1600-0447.2005.00722.x
6. Figueroa-Barra A, Del Aguila D, Cerda M, Gaspar PA, Terissi LD, Durán M, et al. Automatic language analysis identifies and predicts schizophrenia in first-episode of psychosis. *Schizophrenia.* (2022) 8:53. doi: 10.1038/s41537-022-00259-3
7. Tang SX, Hänsel K, Cong Y, Nikzad AH, Mehta A, Cho S, et al. Latent factors of language disturbance and relationships to quantitative speech features. *Schizophr Bull.* (2023) 49:S93–103. doi: 10.1093/schbul/sbac145
8. Roche E, Creed L, MacMahon D, Brennan D, Clarke M. The epidemiology and associated phenomenology of formal thought disorder: A systematic review. *Schizophr Bull.* (2015) 41:951–62. doi: 10.1093/schbul/sbu129
9. Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. *NPJ Schizophr.* (2015) 1:15030. doi: 10.1038/npschz.2015.30
10. Hoekert M, Kahn R, Pijnenborg M, Aleman A. Impaired recognition and expression of emotional prosody in schizophrenia: Review and meta-analysis. *Schizophr Res.* (2007) 96:135–45. doi: 10.1016/j.schres.2007.07.023
11. Lundin N, Blouin A, Cowan H, Moe A, Wastler H, Breitborde N. Identification of psychosis risk and diagnosis of first-episode psychosis: advice for clinicians. *Psychol Res Behav Manage.* (2024) 17:1365–83. doi: 10.2147/PRBM.S423865
12. Morgan SE, Diederer K, Vêrtes PE, Ip SHY, Wang B, Thompson B, et al. Natural Language Processing markers in first episode psychosis and people at clinical high-risk. *Transl Psychiatry.* (2021) 11:630. doi: 10.1038/s41398-021-01722-y
13. Haas SS, Doucet GE, Garg S, Herrera SN, Sarac C, Bilgrami ZR, et al. Linking language features to clinical symptoms and multimodal imaging in individuals at clinical high risk for psychosis. *Eur Psychiatry.* (2020) 63:e72. doi: 10.1192/j.eurpsy.2020.73
14. Lucarini V, Grice M, Cangemi F, Zimmermann JT, Marchesi C, Vogeley K, et al. Speech prosody as a bridge between psychopathology and linguistics: the case of the schizophrenia spectrum. *Front Psychiatry.* (2020) 11–2020. doi: 10.3389/fpsy.2020.531863
15. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. *Speech Commun.* (2015) 71:10–49. doi: 10.1016/j.specom.2015.03.004
16. Low LSA, Maddage NC, Lech M, Sheeber LB, Allen NB. Detection of clinical depression in adolescents' Speech during family interactions. *IEEE Trans BioMed Eng.* (2011) 58:574–86. doi: 10.1109/TBME.2010.2091640
17. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]*. New York, NY, USA: Association for Computing Machinery. (2016) p. 785–94. (KDD '16). doi: 10.1145/2939672.2939785
18. Anticevic A, Murray JD, Barch DM. Bridging levels of understanding in schizophrenia through computational modeling. *Clin Psychol Sci.* (2015) 3:433–59. doi: 10.1177/2167702614562041