Global Congress on Public Health 2025

July 23-24, 2025 | Paris, France

Zhen Zeng

Department of Psychiatry, Chongqing Mental Health Center, Chongqing, 400036, China

From Subjective to Objective: The Rise of Automated Language Analysis in Schizophrenia Detection

Language and speech are essential data sources for diagnosing and treating mental disorders, as they provide valuable insights into the organization and content of thought. Traditional approaches to language analysis largely depend on expert opinions, clinical rating scales, and manual linguistic analyses. However, these methods often lack objectivity, are time-consuming, and may be influenced by individual biases, limiting their scalability and reliability in large-scale clinical practice.

In recent years, with the advancement of artificial intelligence and natural language processing (NLP) techniques, it has become possible to extract linguistic features more efficiently and objectively. Automated analysis of language (AAL) refers to the use of computer programs to process and analyze natural language, whether spoken or written. Unlike traditional human-based assessments, AAL offers high objectivity by minimizing human biases, enables rapid and large-scale analysis, and supports the extraction of multi-level language features. Moreover, it can be applied across different modalities and platforms, making it highly versatile.

The general process of AAL involves collecting language data through free conversations, structured interviews, or written materials. This data is then analyzed to extract semantic, syntactic, and emotional features. When combined with machine learning algorithms, AAL can significantly improve the automatic detection, classification, and monitoring of mental disorders. Such data-driven approaches shift the assessment paradigm from subjective interpretation to objective, evidence-based evaluation, providing a scalable pathway for early identification, subtype diagnosis, and treatment response tracking.

Current studies have examined various linguistic dimensions, including lexical features (e.g., word frequency, vocabulary richness, specific word categories), syntactic features (e.g., sentence length, syntactic complexity, part-of-speech tagging), and semantic features (e.g., semantic coherence, referential cohesion, poverty of content, and metaphorical language). Analytical techniques such as latent semantic analysis (LSA) and part-of-speech tagging analyses have demonstrated high accuracy in distinguishing individuals with schizophrenia from healthy controls, and in predicting outcomes in clinical high-risk (CHR) populations.

With the emergence of advanced language models like ChatGPT, there is increasing potential to develop AI prediction models that integrate multimodal features, including language, speech, facial expressions, and behavioral patterns. These multimodal AI systems are expected to enhance the early detection of schizophrenia, improve understanding of its underlying mechanisms, and facilitate comprehensive characterization of its pathological features. Ultimately, this progress will drive forward the field of digital mental health and support more personalized, precise, and proactive mental health care.

ISBN: 978-1-917892-10-0

doi.org/10.51219/URForum.2025.Zhen-Zeng

Key words: Prodromal phase of schizophrenia, linguistic features, machine learning and natural language analysis, early identification and screening

ISBN: 978-1-917892-10-0

doi.org/10.51219/URForum.2025.Zhen-Zeng