Schizophrenia Research 267 (2024) 65–71

Contents lists available at ScienceDirect

Schizophrenia Research

journal homepage: www.elsevier.com/locate/schres

Automated linguistic analysis in speech samples of Turkish-speaking patients with schizophrenia-spectrum disorders

Berat Arslan [a,*], Elif Kizilay [a], Burcu Verim [a], Cemal Demirlek [a,b], Yagmur Dokuyan [c], Yaren Ecesu Turan [a], Aybuke Kucukakdag [a], Muhammed Demir [a], Ezgi Cesim [a], Emre Bora [a,d,e]

[a] Department of Neurosciences, Health Sciences Institute, Dokuz Eylul University, Izmir, Turkey
[b] Department of Psychiatry, McLean Hospital, Harvard Medical School, Belmont, MA, USA
[c] Department of Psychiatry, Izmir City Hospital, Izmir, Turkey
[d] Department of Psychiatry, Faculty of Medicine, Dokuz Eylul University, Izmir, Turkey
[e] Department of Psychiatry, Melbourne Neuropsychiatry Centre, University of Melbourne and Melbourne Health, Carlton South, Victoria 3053, Australia

ARTICLE INFO

Keywords:
Schizophrenia spectrum disorders
Natural language processing
Semantic similarity
Diagnostic classification
Biomarker

ABSTRACT

Modern natural language processing (NLP) methods provide ways to objectively quantify language disturbances for potential use in diagnostic classification. We performed computerized language analysis in speech samples of 82 Turkish-speaking subjects, including 44 patients with schizophrenia spectrum disorders (SSD) and 38 healthy controls (HC). Exploratory analysis of speech samples involved 16 sentence-level semantic similarity features using SBERT (Sentence Bidirectional Encoder Representation from Text) as well as 8 generic and 8 part-of-speech (POS) features. The random forest classifier using SBERT-derived semantic similarity features achieved a mean accuracy of 85.6 % for the classification of SSD and HC. When semantic similarity features were combined with generic and POS features, the classifier's mean accuracy reached to 86.8 %. Our analysis reflected increased sentence-level semantic similarity scores in SSD. Generic and POS analyses revealed an increase in the use of verbs, proper nouns and pronouns in SSD while our results showed a decrease in the utilization of conjunctions, determiners, and both average and maximum sentence length in SSD compared to HC. Quantitative language features were correlated with the expressive deficit domain of BNSS (Brief Negative Symptom Scale) as well as with the duration of illness. These findings from Turkish-speaking interviews contribute to the growing evidence-based NLP-derived assessments in non-English-speaking patients.

1. Introduction

Patients diagnosed with schizophrenia spectrum disorders (SSD) can show a wide range of behavioral, emotional, psychomotor, cognitive, speech and language disturbances, including impairments at the level of acoustic, phonetic, semantics (meaning) and syntax (grammar) (Bleuler, 1950; Chaika, 1990; Covington et al., 2005; DeLisi, 2001). The importance of speech and language disturbances due to their high predictive and prognostic value was highlighted by several studies (Roche et al., 2015; Wilcox et al., 2012). After a wide range of searches for potential biomarkers in clinical diagnosis classification, modern natural language processing (NLP) methods offer means of quantifying semantic similarity and syntactic features objectively for the possible use of the diagnosis classification (Corcoran et al., 2018; Figueroa-Barra et al., 2022; Tang et al., 2021). These methods can be conducted noninvasively, efficiently and inexpensively, making them potentially useful for diagnosis classification (Tan and Rossell, 2020).

Disturbances at the semantic level in SSD have been investigated using a wide range of NLP methods (Corcoran et al., 2020; Hitczenko et al., 2021; Ratana et al., 2019). The rapid development of novel NLP methods for semantic similarity analysis has led to the use of various methods including latent semantic analysis, semantic density, word2vec and BERT, which is a more recent NLP technique that has consistently outperformed other techniques in a wide range of NLP tasks (Devlin et al., 2019). After the pioneering works of Elvevåg et al. (2010, 2007) for the quantification of semantic similarity using latent semantic analysis, several subsequent studies in native English individuals have been conducted to predict later psychosis development (Bedi et al., 2015; Corcoran et al., 2018). Another study predicted the development of psychosis at follow-up in individuals at ultra-high risk for psychosis

* Corresponding author.
E-mail address: arslan.ufuk.berat@gmail.com (B. Arslan).

https://doi.org/10.1016/j.schres.2024.03.014
Received 10 December 2023; Received in revised form 5 February 2024; Accepted 14 March 2024
Available online 22 March 2024

using semantic density, obtained using vector unpacking (Rezaii et al., 2019). Tang et al. (2021) conducted a sentence-level analysis using BERT to show increased tangentiality in SSD and they reported NLP measures were even better than clinical ratings. Furthermore, Chan et al. (2023) showed an increased expression of topics related to self-experience in schizophrenia and they classified schizophrenia and healthy controls (HC) using automated NLP features, including sentence-level semantic similarity features. In addition to these studies, numerous studies have been conducted on schizophrenia in languages other than English. In native Dutch speakers, patients with SSD were differentiated from HC using the semantic word2vec model by the moving window approach (Voppel et al., 2021). The same research group showed that SSD and HC classification can be improved by a combined machine-learning model using semantic similarity and acoustic features (Voppel et al., 2023). Figueroa-Barra et al. (2022) classified schizophrenia and HC in Spanish-speaking individuals using verbal fluency, verbal productivity and semantic similarity features. Additionally, several studies suggest that language in SSD is susceptible to antipsychotic medication (de Boer et al., 2020a; Sinha et al., 2015).

In addition to the methods mentioned earlier, a novel approach called Sentence Bidirectional Encoder Representation from Text (SBERT) has been developed. SBERT is fine-tuned BERT using Siamese and triplet network architecture and SBERT has also been shown to exhibit a significant improvement compared to the state-of-the-art techniques for generating sentence embeddings (Reimers and Gurevych, 2019). SBERT was employed in a study, that reported higher semantic similarity to the Inventory of Psychotic-Like Anomalous Self-Experiences (IPASE) items in individuals at clinical high-risk for psychosis as compared to both healthy controls and patients with psychosis (Srivastava et al., 2023).

In addition to studies based on semantic similarity, it is important to note that language should not be assessed by focusing solely on semantic aspects given its complexity. Numerous studies using automated methods have been conducted on structural aspects of language, such as syntax, which can provide valuable insights for understanding the extent and severity of language and communication difficulties in patients with SSD (de Boer et al., 2020a). Bedi et al. (2015) revealed significant changes in maximum phrase length and frequency of part-of-speech (POS) tags, particularly in the use of determiners within the clinical high-risk psychosis group. Subsequent studies have continued this exploration in patients with clinical high risk for psychosis, utilizing NLP-derived features such as the number of content words, unique words, sentences, words per sentence and frequency of POS tags (Corcoran et al., 2018; Gupta et al., 2018; Rezaii et al., 2019). Another study conducted POS analysis more in detail and reported increased use of first-person singular pronouns but decreased use of determiners, adverbs and adjectives in SSD compared to HC (Tang et al., 2021). In addition to differences between patients and healthy controls, individuals with schizophrenia did not show a difference in type-token ratio compared to HC, however, they exhibited decreased use of words per sentence and used shorter words (Hong et al., 2015). Following studies based on NLP methods examining the structure of language from various aspects, it was found that the integrity of the white matter in the language tracts and negative symptoms were associated with quantitative measures of language structure, such as mean length of utterance (de Boer et al., 2020b). Further methodologies, including speech graph analysis (Mota et al., 2012, 2017) and Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2015), were employed in several studies. Notably, LIWC was applied to assess correlations between NLP measures and clinical symptoms (Buck and Penn, 2015; Minor et al., 2015).

Despite the notable successes of the novel computational methods, NLP methods are still in an unstandardized and rapidly evolving state. Additionally, most of the studies have primarily focused on participants who are fluent in English. In the current exploratory study, we expect that speech from Turkish-speaking patients with SSD would exhibit abnormalities in NLP-derived language features in their speech. Our first aim is to quantify sentence-level semantic similarity using SBERT and analyze generic and POS features in SSD and HC. Additionally, we will compare these linguistic features between the groups while exploring any potential correlation between the linguistic features and clinical symptoms. Our second aim is to adapt separate machine-learning approaches for semantic similarity as well as generic and POS features to differentiate SSD from HC, assess the accuracy of the classification, and identify the most informative features of the models. Finally, we intend to merge semantic similarity features with generic and POS features to train a comprehensive model to evaluate the classification accuracy and determine the most significant features of the final model.

2. Methods

2.1. Participants

Speech samples were recorded from 44 SSD participants (31 schizophrenia and 13 schizoaffective disorder) and 38 HC at the Department of Neuroscience and Department of Psychiatry, Dokuz Eylul University. Patients were included through the Psychotic Disorders Outpatient Unit of the Department of Psychiatry. HCs were included through advertisements on the Dokuz Eylül University Medical School campus and at the university hospital if they had no personal or familial history of psychiatric disorders. All participants were adult native Turkish speakers.

For all participants, personal history of medical disorders, current alcohol/substance abuse or personal history of neurological disorders were exclusion criteria. All participants were interviewed using the Structured Clinical Interview for DSM-IV Axis I Disorders. Current psychotic, formal thought disorder (FTD), and negative symptoms of all patients were evaluated using the Scale for the Assessment of Positive Symptoms (SAPS) (Andreasen, 1984) and the Brief Negative Symptom Scale (BNSS) (Kirkpatrick et al., 2011; Polat Nazlı et al., 2016; Weigel et al., 2023). Individuals diagnosed with SSD were treated with antipsychotic medication and the dosage of this medication was standardized to chlorpromazine equivalent (Leucht et al., 2014). Each individual signed a written informed consent form and the study was conducted following the approval of the ethics committee of the Dokuz Eylul University.

2.2. Interview procedure

Speech was elicited by a trained researcher from a semi-structured Turkish-speaking interview of approximately 15 min using The Discourse in Psychosis Speech Bank Protocol (https://discourseinpsychosis.org/) which is intended for a multilingual dataset for studying language in psychosis. The Protocol includes free conversational speech, personal narrative, health narrative, picture descriptions, storyboard, dream reports, and reading-recall task sections. Since the section for the health narrative was not collected from every participant and the reading task did not fit the purpose of semantic similarity analysis, the section for the health narrative and the reading part of the reading-recall task section were excluded from the analysis. If a participant failed to initiate speech or had a long-lasting pause, occasional questions were used to encourage the participant. A complete list of questions is available in Table S1.

2.3. Speech pre-processing

Audio files were manually transcribed. Filled pauses were removed during the transcription process. However, repetitions were not eliminated. Transcribers who are native Turkish speakers manually identified sentence boundaries based on syntax and pauses indicated by punctuation marks. Two authors checked the sentence boundaries. The text was preprocessed before the transcripts were analyzed further. Python 3 was used for text preprocessing. After the transcripts were converted to lowercase, the Natural Language Toolkit (NLTK; http://www.nltk.org) was used to parse each transcript into sentences, relying on the punctuation provided by the transcribers to determine the boundaries of sentences, and then all punctuation marks were eliminated using a Spacy Turkish model (Altinok, 2023) trained on the Turkish BOUN Treebank (Türk et al., 2022). To conduct generic and part-of-speech (POS) analyses, words were tokenized and stop words were eliminated using NLTK. For the POS analysis, each word was tagged based on its grammatical function in Turkish using the Spacy Turkish previously mentioned.

2.4. Semantic similarity analysis

The sentence-level semantic similarity analysis was performed without lemmatization or elimination of stop words. The semantic similarity analysis was conducted for each transcript in a series of sentences in lowercase and without punctuation. After the preprocessing steps, sentences were vectorized using SBERT, a modification of the BERT network using Siamese and triplet networks to compute semantically meaningful sentence embeddings (Reimers and Gurevych, 2019, 2020).

The exploratory analysis was performed based on cosine similarity, which is a measurement of vector proximity used to quantify and evaluate semantic similarity between sentences encoded by representative vectors. To conduct the similarity analysis, we utilized a pre-trained model from the Sentence Transformers framework. This framework includes a multi-lingual model of the universal sentence encoder that supports 15 languages, including Turkish (sentence-transformers/distiluse-base-multilingual-cased-v1).

Four cosine similarity-based metrics were used to obtain the semantic similarity features. The first three metrics were adapted from Chan et al. (2023)'s study. We computed the cosine similarity between adjacent sentences (first metric), between a sentence and the sentence spoken 2 sentences later (second metric), between a sentence and the sentence spoken 3 sentences later (third metric), and the similarity of each sentence to each other sentence (fourth metric) resulting in a set of sentence similarity values for each of the 4 metrics. For each set of semantic similarity values, we calculated the mean, variance, 5th percentile and 95th percentile of similarity values resulting in a total of 16 sentence-level semantic similarity features.

2.5. Generic and POS analyses

Eight generic features and eight POS features were carried out. These measurements encompassed various aspects, including the number of total words and sentences, the number of unique words (type-token ratio) and stop words per word, the average length of both words and words without stop words, as well as average and maximum sentence length. Additionally, we captured the number of eight different tags -noun, verb, adverb, proper noun, adjective, pronoun, conjunction, determiner- per word. A complete list of generic and POS features is available in Table 2.

2.6. Machine-learning classification and statistical analyses

The scikit-learn package of Python was used to conduct three random forest analyses for assessing the classification accuracies. A random forest classifier was performed using 16 semantic similarity features. A separate random forest classifier was carried out using 8 generic and 8 POS features. Additionally, a combined model was performed by combining all features used in separate classifiers. All models were performed using 100 decision trees. For each model training, 10-fold cross-validation was used, where 90 % of the data set was used as a training sample and 10 % as a testing sample. 10-fold cross-validation was performed using Stratified KFold Cross Validation ensuring that each fold retains an equal distribution of different groups. For each model, this process was repeated to test different data samples each time and cross-validation splits were the same for each classification. Additionally, mean accuracies were calculated across 10-fold cross-validation. Given the limited sample size in the current study, a holdout dataset was not included.

Statistical analyses were performed using Jamovi 2.3.28 (The jamovi project, 2022). Chi-square tests were performed for categorical variables and independent samples *t-tests* were used for continuous variables to evaluate demographic differences between groups. The Shapiro–Wilk normality test was used for checking normal distribution. The group's statistical differences in the NLP-derived features were assessed using independent *t-tests*. Effect sizes were assessed using Cohen's d. Associations between language features and severity of clinical symptoms were assessed using Pearson correlations. Additionally, false discovery rate (FDR) correction was applied to correlations.

3. Results

3.1. Demographics

Age or gender did not differ significantly between SSD and HC. A significant difference was found between SSD and HC in terms of education level, but no significant difference was found in parental years of education. A full list of demographic and clinical data is available in Table 1.

3.2. Semantic similarity

In the exploratory analysis, compared to the HC group, the SSD group exhibited significantly higher semantic similarity in all four metrics. Increased semantic similarity was found in the first metric, which was computed using the cosine similarity between adjacent sentences, including mean ($t = -6.32$, $d = -1.4$, $p < 0.001$), variance ($t = -6.53$, $d = -1.45$, $p < 0.001$), 5th percentile ($t = -2.3$, $d = -0.5$, $p = 0.024$) and 95th percentile ($t = -7.1$, $d = -1.57$, $p < 0.001$). Similar results were observed in semantic similarity scores in the second and third metrics (all $p < 0.05$). Additionally, increased similarity scores for participants with SSD were found in the fourth metric, which was computed using the similarity of each sentence to each other sentence, including mean ($t = -5.28$, $d = -1.17$, $p < 0.001$), variance ($t = -4.37$, $d = -0.97$, $p < 0.001$), 5th percentile ($t = -4.97$, $d = -1.1$, $p < 0.001$)

Table 1
Demographic and clinical parameters.

	SSD patients ($n = 44$)	Healthy controls ($n = 38$)	Statistics
Age (years)	35.48 (10.29)	35.03 (11.71)	$t = -0.186$, $p = 0.853$
Gender (male/female)	25/19	25/13	$\chi2 = 0.690$, $p = 0.406$
Years of education	11.75 (3.24)	13.79 (3.52)	$t = 2.733$, $p = 0.008$
Years of parental education	7.04 (3.89)	8.35 (4.36)	$t = 1.357$, $p = 0.179$
Age of onset of illness (years)	22.2 (5.97)		
Duration of illness (months)	152 (121)		
CPZ equivalents (mg/day)	646 (392)		
SAPS total	16.1 (16.9)		
SAPS positive FTD	3.44 (5.06)		
BNSS total	21.7 (13.9)		
BNSS (MAP)	14.3 (7.97)		
BNSS (EXP)	7.40 (6.82)		

SAPS: Scale for the Assessment of Positive Symptoms, FTD: Formal Thought Disorder, BNSS: Brief Negative Symptom Scale, MAP: Motivational Deficit Domain of BNSS, EXP: Expressive Deficit Domain of BNSS, CPZ: chlorpromazine.

and 95th percentile ($t = -5.48$, $d = -1.21$, $p < 0.001$). A complete list of semantic similarity results is available in Table S2.

3.3. Generic and POS features

The analysis revealed a decrease in total words used in SSD compared to HC ($p = 0.006$) while the number of total sentences did not differ between SSD and HC. A decrease was revealed in the average length of words after removing stop words in SSD ($p = 0.003$). However, no significant difference was found in the average length of words without removing stop words. Although our analysis revealed a significantly lower number of stop words per word ($p = 0.01$) in SSD, there was no significant difference in the number of unique words per word (type-token ratio). In SSD, a decrease in both average ($p < 0.001$) and maximum ($p < 0.001$) sentence length was found. Moreover, the SSD group showed a higher number of verbs ($p = 0.027$), proper nouns ($p = 0.021$), and pronouns ($p = 0.013$) per word, along with a decrease in number of conjunctions ($p = 0.014$) and determiners ($p = 0.003$) per word. No significant differences were observed in number of nouns, adverbs, or adjectives per word. A complete list of generic and POS results is available in Table 2.

3.4. Correlations with clinical ratings

After false discovery rate (FDR) correction, exploratory correlation analysis showed that duration of illness correlated with 95th percentile ($r = 0.423$, $p = 0.0376$) of the first metric as well as mean ($r = 0.430$, $p = 0.032$) and variance ($r = 0.463$, $p = 0.0144$) of the second metric. Moreover, the expressive deficit domain (EXP) of the BNSS correlated with variance ($r = 0.437$, $p = 0.0272$) and 95th percentile ($r = 0.431$, $p = 0.0312$) of the fourth metric (Fig. 1). Furthermore, the positive FTD domain of the SAPS showed a significant correlation with maximum sentence length ($r = 0.537$, $p = 0.0016$). Additionally, language features were not correlated with the age of onset of illness, chlorpromazine equivalents, SAPS total, BNSS total and the motivational deficit domain (MAP) of the BNSS. A complete list of exploratory correlation results is available in Table S3.

Table 2
Comparison of generic and POS features between SSD and HC.

Feature name	SSD (n = 44)	HC (n = 38)	T-Statistic	*p*-Value	Cohen's d
Number of total sentences	181.48 (59.60)	172.39 (47.85)	−0.753	0.454	−0.17
Number of total words	894.23 (369.98)	1107.18 (296.31)	2.846	0.006**	0.63
Average length of words	5.58 (0.20)	5.66 (0.21)	1.629	0.107	0.36
Average length of words without stop words	6.09 (0.23)	6.24 (0.21)	3.115	0.003**	0.69
Number of stop words per word	0.16 (0.03)	0.17 (0.02)	2.635	0.01*	0.58
Number of unique words per word	0.54 (0.06)	0.54 (0.04)	0.237	0.813	0.05
Average sentence length	4.88 (1.09)	6.59 (1.49)	5.993	<0.001**	1.33
Maximum sentence length	20.82 (7.88)	29.21 (10.19)	4.199	<0.001**	0.93
Number of verbs per word	0.22 (0.02)	0.21 (0.02)	−2.247	0.027*	−0.50
Number of nouns per word	0.30 (0.03)	0.30 (0.03)	−0.993	0.323	−0.22
Number of adverbs per word	0.12 (0.03)	0.12 (0.02)	0.217	0.828	0.05
Number of proper nouns per word	0.006 (0.004)	0.004 (0.003)	−2.354	0.021*	−0.52
Number of adjectives per word	0.10 (0.02)	0.11 (0.02)	0.945	0.348	0.21
Number of pronouns per word	0.06 (0.02)	0.05 (0.01)	−2.546	0.013*	−0.56
Number of conjunctions per word	0.065 (0.019)	0.075 (0.014)	2.523	0.014*	0.56
Number of determiners per word	0.066 (0.015)	0.076 (0.013)	3.115	0.003**	0.69

Means and standard deviations (SD) are listed for each group and generic and POS features. SSD: schizophrenia-spectrum disorders; HC: healthy controls.
* Indicates significance at $p < 0.05$.
** Indicates significance at $p < 0.01$.

3.5. Machine-learning results

The HC vs. SSD classification of the 10-fold cross-validated random forest classifier using 16 semantic similarity features achieved a mean accuracy of 85.6 %, with a 95 % CI [0.771, 0.942]. The area under the curve-receiver operating characteristic (AUC-ROC) of the model using 16 semantic similarity features was 0.89. Upon ranking by feature importance, the top informative feature was the 95th percentile of the first metric.

The 10-fold cross-validated random forest classifier using 8 generic and 8 POS features distinguished patients with SSD from HC with a mean accuracy of 68.4 %, with a 95 % CI [0.594, 0.774]. The AUC-ROC for this classification was 0.76. For the model, the most informative feature was average sentence length. A list of the top five features of both classifiers is available in Fig. 2.

The 10-fold cross-validated combined model using the complete list of features, a total of 32 features, had a mean accuracy of 86.8 %, with a 95 % CI [0.765, 0.970]. The AUC-ROC reached 0.89 in the combined model and the top informative feature was the variance of the first metric. The second-highest-ranked feature was the 95th percentile of the first metric, which was the highest-ranked semantic similarity feature in the model of semantic similarity features only. Out of the top 10 features, nine were from the semantic similarity domain. However, the eighth feature was average sentence length. A list of the top 10 features and the AUC-ROC curve of the combined model are available in Fig. 3.

4. Discussion

The present exploratory study aimed to investigate linguistic abnormalities of Turkish-speaking patients with SSD using SBERT and to assess semantic similarity at the sentence level, alongside generic and POS analyses. Furthermore, we sought to explore the relationship between language characteristics and clinical symptoms while evaluating the robustness of the machine-learning approach for the classification of patients with SSD and HC.

Regarding semantic similarity analysis, sentence-level SBERT measures indicated a greater variance and a higher degree of semantic similarity in patients with SSD. In our findings, there was an increase in mean, 5th percentile, and 95th percentile of sentence-level semantic similarity scores. Our finding of an increase in variance of similarity scores at the sentence level aligns with previous research (Corcoran et al., 2018; Voppel et al., 2021). Voppel et al. (2021) found an increased variance in SSD within sentence-length windows. Furthermore, several studies have indicated a decrease in semantic similarity scores among individuals with psychosis (Bedi et al., 2015; Elvevåg et al., 2007; Lundin et al., 2022). Conversely, a recent study revealed that patients with first-episode schizophrenia had higher semantic similarity scores (Alonso-Sánchez et al., 2023), which is consistent with our finding of an increase in semantic similarity scores in patients with SSD. In addition, Palominos et al. (2023) performed a preliminary analysis of semantic similarity based on fastText, resulting in no significant differences in semantic similarity between schizophrenia, clinical high risk and HC. Taken together, these contradictory findings raise the question of

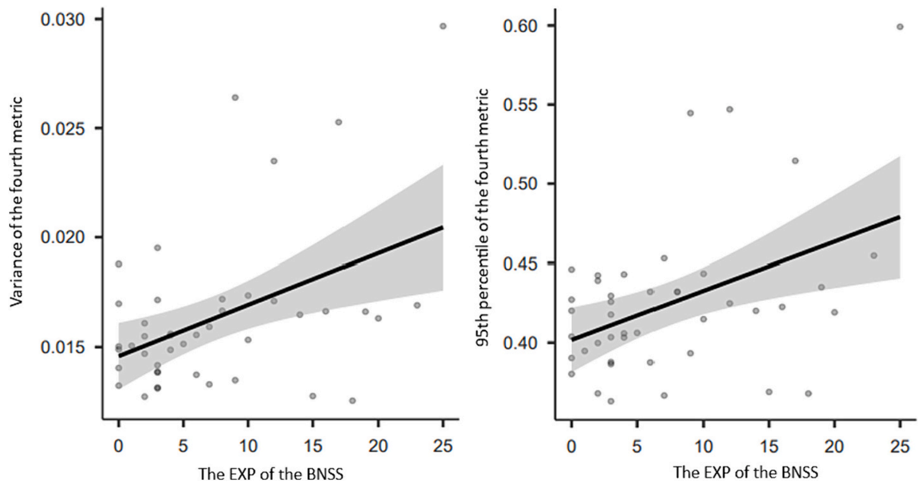

Fig. 1. The expressive deficit domain (EXP) of the BNSS was positively correlated with variance (r = 0.437) (left) and 95th percentile (r = 0.431) (right) of the fourth metric.

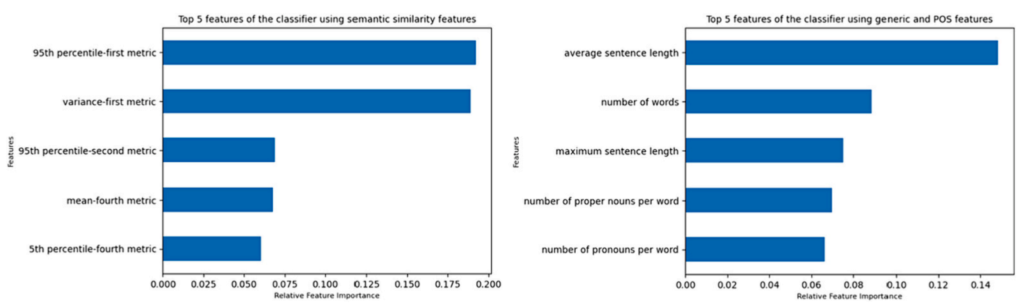

Fig. 2. Left. Relative feature importance for the classifier using semantic similarity features. Right. Relative feature importance for the classifier using generic and POS features.

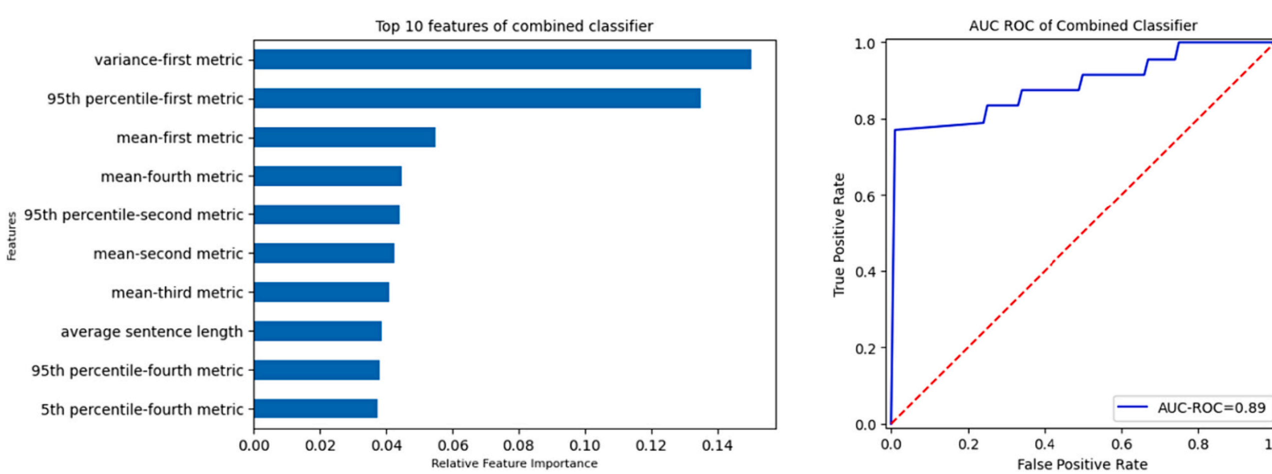

Fig. 3. Left. Relative feature importance for the combined classifier. Right. AUC-ROC curve of the combined classifier.

whether the idea of semantic similarity measurement is a valid method for assessing coherence using semantic NLP methods.

Concerning the generic and POS measures, we found an increased use of pronouns in SSD, which aligns with findings from multiple studies that have reported an increase in pronoun usage among SSD (Buck and Penn, 2015; Tang et al., 2021). Determiners have been examined and used to classify patients in several studies (Bedi et al., 2015; Tang et al., 2021). In our study, patients with SSD showed a decrease in their use of determiners. Moreover, we observed significant differences in the generic measures. Our finding of a decrease in average sentence length in SSD is in line with those of several other studies (Elvevåg et al., 2010; Hong et al., 2015; Jeong et al., 2023). Furthermore, our finding of a decrease in the average length of words without stop words aligns with those of prior studies (Figueroa-Barra et al., 2022; Jeong et al., 2023). However, our findings did not demonstrate a significant difference in the type-token ratio, which has been identified as a differentiating factor between groups in previous research (Figueroa-Barra et al., 2022). Based on our findings, the absence of a difference in the number of total sentences, combined with our findings indicating a decrease in overall word count, average sentence length and use of stop words in SSD, suggest that individuals with SSD tended to speak less and used simpler sentences compared to HC. Additionally, our analysis revealed a higher frequency of verbs and proper nouns in individuals with SSD, while the use of conjunctions was found to be lower.

Our exploratory correlation analysis revealed that maximum sentence length positively correlated with the positive FTD domain of the SAPS. Concerning negative symptoms, two features were significantly correlated with the EXP constituted by blunted affect and alogia. Variance and 95th percentile of the fourth semantic similarity metric, the similarity of each sentence to each other sentence, were positively correlated with the EXP, suggesting that variance and 95th percentile of the fourth semantic similarity metric could potentially be a significant indicator for the severity of blunted affect and alogia in SSD. Furthermore, our results indicate increased similarity scores of all semantic similarity metrics in SSD, suggesting that the sentences produced by patients with SSD diverge less from prior sentences. This points to a more restricted semantic space in the speech of these patients. The semantic distance between sentences of patients with SSD is narrow, which might indicate the poverty of thought in patients with SSD. This was supported by our correlation findings between the fourth semantic similarity metric, and blunted affect and alogia. Additionally, 95th percentile of the first metric, alongside mean and variance of the second metric, positively correlated with duration of illness.

In the current study, our classifier, using SBERT-derived semantic similarity features, achieved a mean accuracy of 85.6 %. High accuracy has been achieved using semantic similarity features alone, without combining additional features such as acoustic or syntactic parameters, indicating that the robustness of SBERT for generating sentence embeddings might be a contributing factor behind this outcome (Reimers and Gurevych, 2019). Concerning the classifier using generic and POS features, the mean accuracy was 68.4 %. Moreover, the classifier reached a mean accuracy of 86.8 %, after combining semantic similarity features with generic and POS features. The combined classifier achieved a similar accuracy to the classifier using semantic similarity features alone. The results of classifiers indicate that semantic similarity analysis has greater discriminative ability between groups compared to generic and POS analyses. This could also mean that sentence-level semantic embeddings derived using SBERT may capture differences in generic and POS features between SSD and HC. In addition, our accuracy in classifying Turkish speech samples was similar to that of other languages, providing further support for the utilization of automated semantic features reflecting thought disorder in schizophrenia across different languages (Figueroa-Barra et al., 2022; Iter et al., 2018; Tang et al., 2021). Here, we showed that classification based on automated linguistic features can be a promising tool for Turkish-speaking individuals. Furthermore, several studies performed classifiers based on serum, EEG, MRI and PET data to assess accuracy levels (Kambeitz et al., 2015; Schwarz et al., 2010; Shim et al., 2016; Zeng et al., 2018), which can be comparable to those of our classifier. In addition to these factors, the NLP method has been shown to outperform classifications using clinical measures (Bedi et al., 2015; Tang et al., 2021). Given these indicators, automated language analysis could potentially be employed as a tool for diagnostic classification.

This study has several limitations. Although there were no significant differences in age, gender and years of parental education, it is important to note that the groups had different education levels, which may have introduced a potential bias into our findings. Although our machine-learning model showed high accuracy in classifying patients with SSD and HC, given the limited sample size, there is a potential for overfitting. In addition, to prevent overfitting, our classifiers employed 10-fold cross-validation. However, a holdout dataset was not available in our study.

5. Conclusion

In conclusion, we demonstrated that automated linguistic analysis shows potential as a method to quantify language disturbances in SSD, to assess associations between these disturbances and clinical symptoms and to classify patients with SSD and HC with high accuracy based on semantic similarity analysis, which showed similar discriminative ability to the combined classifier. To the best of our knowledge, this is the first study using NLP-derived measures to characterize language disturbances in Turkish-speaking patients. Our findings add to the increasing body of evidence on non-English-speaking patients using NLP-derived measures. Refining these measures can help unlock the potential of automated language analysis to yield clinical biomarkers for noninvasive, efficient, and affordable patient classification as well as characterization of language disturbances in psychiatric disorders.

Role of the funding source

This study was supported by the Scientific and Technological Research Council of Turkey (TUBITAK 2247 Project No: 120C141) which had no further role in the study design, in the collection, analysis and interpretation of data and in the decision to submit the article for publication.

CRediT authorship contribution statement

Berat Arslan: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Elif Kizilay:** Writing – review & editing, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Burcu Verim:** Investigation. **Cemal Demirlek:** Investigation. **Yagmur Dokuyan:** Investigation. **Yaren Ecesu Turan:** Investigation. **Aybuke Kucukakdag:** Investigation. **Muhammed Demir:** Investigation. **Ezgi Cesim:** Investigation. **Emre Bora:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors have no conflicts of interest regarding subject of this manuscript.

Acknowledgement

Not relevant.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.schres.2024.03.014.

References

Alonso-Sánchez, M.F., Limongi, R., Gati, J., Palaniyappan, L., 2023. Language network self-inhibition and semantic similarity in first-episode schizophrenia: a computational-linguistic and effective connectivity approach. Schizophr. Res. 259, 97–103. https://doi.org/10.1016/j.schres.2022.04.007.

Altinok, D., 2023. A diverse set of freely available linguistic resources for Turkish. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Assoc. Comput. Linguist, pp. 13739–13750. https://doi.org/10.18653/v1/2023.acl-long.768.

Andreasen, N.C., 1984. The Scale for the Assessment of Positive Symptoms (SAPS). The University of Iowa Press, Iowa City, IA.

Bedi, G., Carrillo, F., Cecchi, G.A., Slezak, D.F., Sigman, M., Mota, N.B., Ribeiro, S., Javitt, D.C., Copelli, M., Corcoran, C.M., 2015. Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophr. 1, 15030 https://doi.org/10.1038/npjschz.2015.30.

Bleuler, E., 1950. Dementia Praecox or the Group of Schizophrenias. International Universities Press.

Buck, B., Penn, D.L., 2015. Lexical characteristics of emotional narratives in schizophrenia: relationships with symptoms, functioning, and social cognition. J. Nerv. Ment. Des. 203 (9), 702–708. https://doi.org/10.1097/NMD.0000000000000354.

Chaika, E.O., 1990. Understanding Psychotic Speech: Beyond Freud and Chomsky. Charles C. Thomas, Springfield, Illinois.

Chan, C.C., Norel, R., Agurto, C., Lysaker, P.H., Myers, E.J., Hazlett, E.A., Corcoran, C.M., Minor, K.S., Cecchi, G.A., 2023. Emergence of language related to self-experience and agency in autobiographical narratives of individuals with schizophrenia. Schizophr. Bull. 49 (2), 444–453. https://doi.org/10.1093/schbul/sbac126.

Corcoran, C.M., Carrillo, F., Fern Andez-Slezak, D., Bedi, G., Klim, C., Javitt, D.C., Bearden, C.E., Cecchi, G.A., 2018. Prediction of psychosis across protocols and risk cohorts using automated language analysis. World Psychiatry 17 (1), 67–75.

Corcoran, C.M., Mittal, V.A., Bearden, C.E., Gur, R.E., Hitczenko, K., Bilgrami, Z., Savic, A., Cecchi, G.A., Wolff, P., 2020. Language as a biomarker for psychosis: a natural language processing approach. Schizophr. Res. 226, 158–166. https://doi.org/10.1016/j.schres.2020.04.032.

Covington, M.A., He, C., Brown, C., Naçi, L., McClain, J.T., Fjordbak, B.S., Semple, J., Brown, J., 2005. Schizophrenia and the structure of language: the linguist's view. Schizophr. Res. 77, 85–98. https://doi.org/10.1016/j.schres.2005.01.016.

de Boer, J.N., Brederoo, S.G., Voppel, A.E., Sommer, I.E.C., 2020a. Anomalies in language as a biomarker for schizophrenia. Curr. Opin. Psychiatry 33 (3), 212–218. https://doi.org/10.1097/YCO.0000000000000595.

de Boer, J.N., van Hoogdalem, M., Mandl, R.C.W., Brummelman, J., Voppel, A.E., Begemann, M.J.H., van Dellen, E., Wijnen, F.N.K., Sommer, I.E.C., 2020b. Language in schizophrenia: relation with diagnosis, symptomatology and white matter tracts. NPJ Schizophr. 6, 10. https://doi.org/10.1038/s41537-020-0099-3.

DeLisi, L.E., 2001. Speech disorder in schizophrenia: review of the literature and exploration of its relation to the uniquely human capacity for language. Schizophr. Bull. 27 (3), 481–496. https://doi.org/10.1093/oxfordjournals.schbul.a006889.

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In: NAACL HLT 2019–2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.1, pp. 4171–4186. https://doi.org/10.48550/arXiv.1810.04805.

Elvevåg, B., Foltz, P.W., Weinberger, D.R., Goldberg, T.E., 2007. Quantifying incoherence in speech: an automated methodology and novel application to schizophrenia. Schizophr. Res. 93 (1–3), 304–316. https://doi.org/10.1016/j.schres.2007.03.001.

Elvevåg, B., Foltz, P.W., Rosenstein, M., DeLisi, L.E., 2010. An automated method to analyze language use in patients with schizophrenia and their first-degree relatives. J. Neurolinguistics 23 (3), 270–284. https://doi.org/10.1016/j.jneuroling.2009.05.002.

Figueroa-Barra, A., Del Aguila, D., Cerda, M., Gaspar, P.A., Terissi, L.D., Durán, M., Valderrama, C., 2022. Automatic language analysis identifies and predicts schizophrenia in first-episode of psychosis. Schizophrenia 8, 53. https://doi.org/10.1038/s41537-022-00259-3.

Gupta, T., Hespos, S.J., Horton, W.S., Mittal, V.A., 2018. Automated analysis of written narratives reveals abnormalities in referential cohesion in youth at ultra high risk for psychosis. Schizophr. Res. 192, 82–88. https://doi.org/10.1016/j.schres.2017.04.025.

Hitczenko, K., Mittal, V.A., Goldrick, M., 2021. Understanding language abnormalities and associated clinical markers in psychosis: the promise of computational methods. Schizophr. Bull. 47 (2), 344–362. https://doi.org/10.1093/schbul/sbaa141.

Hong, K., Nenkova, A., March, M.E., Parker, A.P., Verma, R., Kohler, C.G., 2015. Lexical use in emotional autobiographical narratives of persons with schizophrenia and healthy controls. Psychiatry Res. 225 (1–2), 40–49. https://doi.org/10.1016/j.psychres.2014.10.002.

Iter, D., Yoon, J.H., Jurafsky, D., 2018. Automatic detection of incoherent speech for diagnosing schizophrenia. In: Proc. 5th Work. Comput. Linguist. Clin. Psychol. From Keyboard to Clin. CLPsych 2018 2018 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol, pp. 136–146. https://doi.org/10.18653/v1/w18-0615.

Jeong, L., Lee, M., Eyre, B., Balagopalan, A., Rudzicz, F., Gabilondo, C., 2023. Exploring the use of natural language processing for objective assessment of disorganized speech in schizophrenia. Psych. Res. Clin. Pract. 5, 84–92. https://doi.org/10.1176/appi.prcp.20230003.

Kambeitz, J., Kambeitz-Ilankovic, L., Leucht, S., Wood, S., Davatzikos, C., Malchow, B., Falkai, P., Koutsouleris, N., 2015. Detecting neuroimaging biomarkers for schizophrenia: a meta-analysis of multivariate pattern recognition studies. Neuropsychopharmacology 40, 1742–1751. https://doi.org/10.1038/npp.2015.22.

Kirkpatrick, B., Strauss, G.P., Nguyen, L., Fischer, B.A., Daniel, D.G., Cienfuegos, A., Marder, S.R., 2011. The brief negative symptom scale: psychometric properties. Schizophr. Bull. 37 (2), 300–305. https://doi.org/10.1093/schbul/sbq059.

Leucht, S., Samara, M., Heres, S., Patel, M.X., Woods, S.W., Davis, J.M., 2014. Dose equivalents for second-generation antipsychotics: the minimum effective dose method. Schizophr. Bull. 40 (2), 314–326. https://doi.org/10.1093/schbul/sbu001.

Lundin, N.B., Jones, M.N., Myers, E.J., Breier, A., Minor, K.S., 2022. Semantic and phonetic similarity of verbal fluency responses in early-stage psychosis. Psychiatry Res. 309, 114404 https://doi.org/10.1016/j.psychres.2022.114404.

Minor, K.S., Bonfils, K.A., Luther, L., Firmin, R.L., Kukla, M., MacLain, V.R., Buck, B., Lysaker, P.H., Salyers, M.P., 2015. Lexical analysis in schizophrenia: how emotion and social word use informs our understanding of clinical presentation. J. Psychiatr. Res. 64, 74–78. https://doi.org/10.1016/j.jpsychires.2015.02.024.

Mota, N.B., Vasconcelos, N.A.P., Lemos, N., Pieretti, A.C., Kinouchi, O., Cecchi, G.A., Copelli, M., Ribeiro, S., 2012. Speech graphs provide a quantitative measure of thought disorder in psychosis. PloS One 7 (4), e34928. https://doi.org/10.1371/journal.pone.0034928.

Mota, N.B., Copelli, M., Ribeiro, S., 2017. Thought disorder measured as random speech structure classifies negative symptoms and schizophrenia diagnosis 6 months in advance. NPJ Schizophr. 3, 18. https://doi.org/10.1038/s41537-017-0019-3.

Palominos, C., Figueroa-Barra, A., Hinzen, W., 2023. Coreference delays in psychotic discourse: widening the temporal window. Schizophr. Bull. 49 (Suppl_2), S153–S162. https://doi.org/10.1093/schbul/sbac102.

Pennebaker, J.W., Boyd, R.L., Jordan, K., Blackburn, K., 2015. The Development and Psychometric Properties of LIWC2015. University of Texas at Austin, Austin, TX. https://doi.org/10.15781/T29G6Z.

Polat Nazlı, I., Ergül, C., Aydemir, Ö., Chandhoke, S., Üçok, A., Gönül, A.S., 2016. Validation of Turkish version of brief negative symptom scale. Int. J. Psychiatry Clin. Pract. 20 (4), 265–271. https://doi.org/10.1080/13651501.2016.1207086.

Ratana, R., Sharifzadeh, H., Krishnan, J., Pang, S., 2019. A comprehensive review of computational methods for automatic prediction of schizophrenia with insight into indigenous populations. Front. Psych. 10, 659. https://doi.org/10.3389/fpsyt.2019.00659.

Reimers, N., Gurevych, I., 2019. Sentence-BERT: sentence embeddings using Siamese BERT-networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pp. 3982–3992. https://doi.org/10.48550/arXiv.1908.10084.

Reimers, N., Gurevych, I., 2020. Making monolingual sentence embeddings multilingual using knowledge distillation. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pp. 4512–4525. https://doi.org/10.48550/arXiv.2004.09813.

Rezaii, N., Walker, E., Wolff, P., 2019. A machine learning approach to predicting psychosis using semantic density and latent content analysis. NPJ Schizophr. 5, 9. https://doi.org/10.1038/s41537-019-0077-9.

Roche, E., Creed, L., Macmahon, D., Brennan, D., Clarke, M., 2015. The epidemiology and associated phenomenology of formal thought disorder: a systematic review. Schizophr. Bull. 41 (4), 951–962. https://doi.org/10.1093/schbul/sbu129.

Schwarz, E., Izmailov, R., Spain, M., Barnes, A., Mapes, J.P., Guest, P.C., Rahmoune, H., Pietsch, S., Markus Leweke, F., Rothermundt, M., Steiner, J., Koethe, D., Kranaster, L., Ohrmann, P., Suslow, T., Levin, Y., Bogerts, B., van Beveren, N., McAllister, G., Weber, N., Niebuhr, D., Cowan, D., Yolken, R.H., Bahn, S., 2010. Validation of a blood-based laboratory test to aid in the confirmation of a diagnosis of schizophrenia. Biomark. Insights 12 (5), 39–47. https://doi.org/10.4137/bmi.s4877.

Shim, M., Hwang, H.J., Kim, D.W., Lee, S.H., Im, C.H., 2016. Machine-learning-based diagnosis of schizophrenia using combined sensor-level and source-level EEG features. Schizophr. Res. 176 (2–3), 314–319. https://doi.org/10.1016/j.schres.2016.05.007.

Sinha, P., Vandana, V.P., Lewis, N.V., Jayaram, M., Enderby, P., 2015. Evaluating the effect of risperidone on speech: a cross-sectional study. Asian J. Psychiatr. 15, 51–55. https://doi.org/10.1016/j.ajp.2015.05.005.

Srivastava, A., Selloni, A., Bilgrami, Z.R., Sarac, C., McGowan, A., Cotter, M., Bayer, J., Spark, J., Krcmar, M., Formica, M., Gwyther, K., Hartmann, J., Ellenberg, E., Polari, A., McGorry, P., Shah, J.L., Yung, A.R., Mizrahi, R., Corcoran, C.M., Cecchi, G.A., Nelson, B., 2023. Differential expression of anomalous self-experiences in spontaneous speech in clinical high-risk and early-course psychosis quantified by natural language processing. Biol. Psychiatry Cogn. Neurosci. Neuroimaging 8 (10), 1005–1012. https://doi.org/10.1016/j.bpsc.2023.06.007.

Tan, E.J., Rossell, S.L., 2020. Questioning the status of aberrant speech patterns as psychiatric symptoms. Br. J. Psychiatry 217 (3), 469–470. https://doi.org/10.1192/bjp.2020.60.

Tang, S.X., Kriz, R., Cho, S., Park, S.J., Harowitz, J., Gur, R.E., Bhati, M.T., Wolf, D.H., Sedoc, J., Liberman, M.Y., 2021. Natural language processing methods are sensitive to sub-clinical linguistic differences in schizophrenia spectrum disorders. NPJ Schizophr. 7, 25. https://doi.org/10.1038/s41537-021-00154-3.

The jamovi project, 2022. jamovi. (Version 2.3) [computer software] (Retrieved from). https://www.jamovi.org.

Türk, U., Atmaca, F., Özateş, Ş.B., Berk, G., Bedir, S.T., Köksal, A., Başaran, B.Ö., Güngör, T., Özgür, A., 2022. Resources for Turkish dependency parsing: introducing the BOUN treebank and the BoAT annotation tool. Lang. Resour. Eval. 56, 259–307. https://doi.org/10.1007/s10579-021-09558-0.

Voppel, A.E., de Boer, J.N., Brederoo, S.G., Schnack, H.G., Sommer, I.E.C., 2021. Quantified language connectedness in schizophrenia-spectrum disorders. Psychiatry Res. 304, 114130 https://doi.org/10.1016/j.psychres.2021.114130.

Voppel, A.E., de Boer, J.N., Brederoo, S.G., Schnack, H.G., Sommer, I.E.C., 2023. Semantic and acoustic markers in schizophrenia-spectrum disorders: a combinatory machine learning approach. Schizophr. Bull. 49 (Suppl_2), S163–S171. https://doi.org/10.1093/schbul/sbac142.

Weigel, L., Wehr, S., Galderisi, S., Mucci, A., Davis, J., Giordano, G.M., Leucht, S., 2023. The brief negative symptom scale (BNSS): a systematic review of measurement properties. Schizophrenia 9, 45. https://doi.org/10.1038/s41537-023-00380-x.

Wilcox, J., Winokur, G., Tsuang, M., 2012. Predictive value of thought disorder in new-onset psychosis. Compr. Psychiatry 53 (6), 674–678. https://doi.org/10.1016/j.comppsych.2011.12.002.

Zeng, L.L., Wang, H., Hu, P., Yang, B., Pu, W., Shen, H., Chen, X., Liu, Z., Yin, H., Tan, Q., Wang, K., Hu, D., 2018. Multi-site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity MRI. EBioMedicine 30, 74–85. https://doi.org/10.1016/j.ebiom.2018.03.017.