Neuroscience Institute Publications   Western Institute for Neuroscience

8-5-2022

Widespread cortical thinning, excessive glutamate and impaired linguistic functioning in schizophrenia: A cluster analytic approach

Liangbing Liang
Western University

Angélica M. Silva
Robarts Research Institute

Peter Jeon
Western University

Sabrina D. Ford
Robarts Research Institute

Michael MacKinley
Robarts Research Institute

See next page for additional authors

Follow this and additional works at: https://ir.lib.uwo.ca/neurosci_inst_pubs

Authors

Liangbing Liang, Angélica M. Silva, Peter Jeon, Sabrina D. Ford, Michael MacKinley, Jean Théberge, and Lena Palaniyappan

frontiers | Frontiers in Human Neuroscience

Check for updates

Widespread cortical thinning, excessive glutamate and impaired linguistic functioning in schizophrenia: A cluster analytic approach

Liangbing Liang[1,2†], Angélica M. Silva[2†], Peter Jeon[3], Sabrina D. Ford[2,4], Michael MacKinley[2,5], Jean Théberge[3,5,6] and Lena Palaniyappan[2,3,5,6,7]*

[1]Graduate Program in Neuroscience, Western University, London, ON, Canada, [2]Robarts Research Institute, Western University, London, ON, Canada, [3]Department of Medical Biophysics, Western University, London, ON, Canada, [4]London Health Sciences Centre, Victoria Hospital, London, ON, Canada, [5]Lawson Health Research Institute, London, ON, Canada, [6]Department of Psychiatry, Western University, London, ON, Canada, [7]Department of Psychiatry, Douglas Mental Health University Institute, McGill University, Montreal, QC, Canada

Introduction: Symptoms of schizophrenia are closely related to aberrant language comprehension and production. Macroscopic brain changes seen in some patients with schizophrenia are suspected to relate to impaired language production, but this is yet to be reliably characterized. Since heterogeneity in language dysfunctions, as well as brain structure, is suspected in schizophrenia, we aimed to first seek patient subgroups with different neurobiological signatures and then quantify linguistic indices that capture the symptoms of "negative formal thought disorder" (i.e., fluency, cohesion, and complexity of language production).

Methods: Atlas-based cortical thickness values (obtained with a 7T MRI scanner) of 66 patients with first-episode psychosis and 36 healthy controls were analyzed with hierarchical clustering algorithms to produce neuroanatomical subtypes. We then examined the generated subtypes and investigated the quantitative differences in MRS-based glutamate levels [in the dorsal anterior cingulate cortex (dACC)] as well as in three aspects of language production features: fluency, syntactic complexity, and lexical cohesion.

Results: Two neuroanatomical subtypes among patients were observed, one with near-normal cortical thickness patterns while the other with widespread cortical thinning. Compared to the subgroup of patients with relatively normal cortical thickness patterns, the subgroup with widespread cortical thinning was older, with higher glutamate concentration in dACC and produced speech with reduced mean length of T-units (complexity) and lower repeats of content words (lexical cohesion), despite being equally fluent (number of words).

Conclusion: We characterized a patient subgroup with thinner cortex in first-episode psychosis. This subgroup, identifiable through macroscopic changes, is also distinguishable in terms of neurochemistry (frontal glutamate) and language behavior (complexity and cohesion of speech). This study supports the hypothesis that glutamate-mediated cortical thinning may contribute to a phenotype that is detectable using the tools of computational linguistics in schizophrenia.

Introduction

Schizophrenia is a disorder that affects how language is employed in everyday use during social interactions (Covington et al., 2005; Kuperberg, 2010; Wible, 2012). Based on the Diagnostic and Statistical Manual of Mental Disorders 5th edition (DSM-5) (American Psychiatric Association, 2013), all of the 5 symptom criteria for diagnosing schizophrenia involve speech and language in one form or another (American Psychiatric Association, 2013). For example, hallucinations are often voices that speak (Alderson-Day et al., 2021); negative symptoms are characterized by "alogia" or reduced speech fluency; thought disorder is expressed as deviations in speech; catatonic features often include mutism (lack of speech production) (Oyebode, 2021); delusions often include an element of misinterpretation of social conversations or deficits in the use of propositional language (Zimmerer et al., 2017). Despite this strong linguistic dependency of the construct of schizophrenia, not every patient diagnosed with this illness displays a detectable speech disturbance (Roche et al., 2015; Kircher et al., 2018; Oomen et al., 2022). It is important to identify patients who are most likely to be afflicted in the language domain, as speech disturbances directly affect the educational and occupational success (Palaniyappan et al., 2019), interpersonal (Tan et al., 2014) and social functioning (Marggraf et al., 2020) as well as endured stigma (Penn et al., 2000). Identification of this subgroup may assist in prognostication in schizophrenia, as well as making early and targeted interventions for a group that may have higher educational and vocational needs possible, before they manifest significant deficits in these domains.

The heterogeneity of linguistic deficits may stem from the presence of a subgroup of patients who do not display the expected language anomalies (Oomen et al., 2022). Alternatively, conventional measures of "formal thought disorder (FTD)" that seek to examine overt communication difficulties may miss the subtle aspects of this deficit, thus introducing an apparent heterogeneity (Mikesell and Bromley, 2016). We need sensitive and objective measures of language indices to study this issue in detail (see Elvevåg et al., 2010; Foltz et al., 2016; Holmlund et al., 2020 for more explanations). One of these tools is natural language processing (NLP) in computational linguistics (Ratana et al., 2019; Corcoran and Cecchi, 2020; Corcoran et al., 2020; Hitczenko et al., 2021). NLP tools use computer algorithms to understand and analyze written text or speech. NLP is a branch of artificial intelligence that uses real-world language as input, and processes it using linguistic rules or patterns identified through statistics, to allow machines to make sense of our language. Such NLP tools do not rely on a clinician's inferential skill to assess the cognitive-linguistic health status (Voleti et al., 2020) of patients from early stages of psychosis (Delvecchio et al., 2019) and are able to predict psychosis onset in individuals at clinical high-risk (CHR) (Bedi et al., 2015). These approaches have broadly focused on syntactic (Thomas et al., 1990; Thomas, 1996; Covington et al., 2005; Delvecchio et al., 2019) and semantic indices (Corcoran et al., 2018; Bar et al., 2019; Alonso-Sánchez et al., 2022; Parola et al., 2022) as the affected domains in psychosis.

Prior studies that focused on quantitative analysis of language have established the following dysfunctions in patients with schizophrenia. First, patients display syntactic simplification [(Morice and Ingram, 1982, 1983; Fraser et al., 1986; Morice and McNicol, 1986; King et al., 1990; DeLisi, 2001; Bilgrami et al., 2022) i.e., they use simple constructions with minimal clause dependencies and also with a limited richness of content]. Secondly, patients show patterns of reduced cohesion (Crider, 1997), for example, lacking prior reference when invoking a description (Chaika and Lambe, 1989) or insufficient lexical repetitions (Gupta et al., 2018) needed to generate cohesion during a discursive discourse (Crossley et al., 2016). Reduced syntactic complexity and cohesion can lead to aberrant word graphs (Mota et al., 2012) and a reduction in number of words spoken (reduced fluency) (Allen et al., 1993; DeLisi, 2001; Morgan et al., 2021).

While some of these features have been linked to the presence of clinically detected FTD, the rating-scale measures of FTD have been poor predictors of linguistic dysfunction *per se* (Mackinley et al., 2021; Tang et al., 2021). Furthermore, as symptom measures fluctuate over time (state-like), they have limited utility in identifying stable subgroups (Jablensky, 2006). Even among speech characteristics, those that relate to "positive symptoms" appear to be more state-related, while those relating to negative symptoms [or Impoverishment of Thinking (Liddle et al., 2002)] appear to be more pervasive. More trait-like measures, e.g., those derived from brain anatomy or genetic composition, that map on to emerging biological insights [e.g., implicating the glutamatergic synapses (Iyegbe and O'Reilly, 2022; Trubetskoy et al., 2022)], may be required to see if specific subgroups of patients have linguistic deficits. Furthermore, as antipsychotics themselves can induce language

impairment (de Boer et al., 2020), recruitment of patients with first-episode psychosis with minimal exposure to antipsychotic medications is necessary to identify subgroups with language dysfunction from illness onset.

In the current study, we first identify subgroups of patients with first-episode schizophrenia using the neuroanatomical measure of MRI-derived cortical thickness. Structural neuroanatomical features are considered to be more stable than symptom rating and physiological recordings, which can vary on a day-to-day basis. In addition, MRI-derived thickness is quantified objectively in an automatized manner with minimal manual intervention in the quantification process. Thus, brain structure can provide more stable and reliable clustering solutions. Further, aberrant cortical thickness has been reported in various illness stages of schizophrenia (Zhao et al., 2022), and has been found to track the inter-individual

TABLE 1   Demographic, clinical, neurobiological, and linguistic data of patients with first-episode psychosis and healthy controls.

	FEP	HC	Pearson's chi-squared test or Welch t-tests	
N	66	36	−	
Demographics				
Age (years)	22.82 (4.77)	21.53 (3.32)	$t(94.043) = 1.6005, p = 0.1128$	
Female/male	12/54	12/24	X-squared (1) = 2.1896, $p = 0.1389$	
Education scale (1/2/3/4)	15/18/20/13	5/3/14/13	X-squared (3) = 8.0131, $p = 0.04574$	*
Clinical				
PANSS-8 (total)	25.18 (6.72)	−	−	
PANSS-8 positive	11.62 (3.48)	−	−	
PANSS-8 negative	6.97 (4.41)	−	−	
PANSS-8 general	5.18 (2.46)			
DUP (weeks) [median (IQR)]	11.0 (4, 24)	−	−	
DDD lifetime exposure [median (IQR)]	0.5 (0, 2.99)	−	−	
Antipsychotic naïve (%)	42%			
Functional				
SOFAS	40.96 (12.40)	−		
NEET status: yes/no	24/29	0/31	X-squared (1) = 17.497, $p < 0.0001$	***
Neurobiological				
Glutamate (mM)	6.79 (1.16)	6.51 (1.35)	$t(53.766) = 0.99493, p = 0.3242$	
Mean cortical thickness (mm)	2.45 (0.12)	2.48 (0.096)	$t(94) = 1.90350, p = 0.0600$	
Language variables				
TLI (Total)	1.48 (1.41)	0.29 (0.39)	$t(81.668) = 6.4188, p < 0.00001$	***
TLI impoverishment	0.57 (0.72)	0.14 (0.25)	$t(87.397) = 4.3669, p < 0.0001$	***
TLI disorganization	0.91 (1.21)	0.15 (0.26)	$t(75.114) = 4.9033, p < 0.00001$	***
Average total number of words	119.18 (38.85)	141.34 (29.83)	$t(88.706) = -3.1775, p = 0.002045$	**
MLS	14.37 (4.58)	14.21 (2.74)	$t(96.753) = 0.20899, p = 0.8349$	
MLT	12.21 (3.00)	12.49 (2.08)	$t(93.295) = -0.56025, p = 0.5767$	
MLC	7.73 (1.20)	8.19 (1.18)	$t(73.659) = -1.8858, p = 0.06327$	
Repeated contents lemmas	0.229 (0.047)	0.247 (0.033)	$t(89.792) = -2.1269, p = 0.03617$	*

Values are reported as "Mean (SD)" unless specified otherwise.

IQR, Interquartile range; FEP, first episode psychosis; HC, healthy controls; PANSS, Positive and Negative Symptoms Scale; DUP, duration of untreated psychosis; DDD, Defined Daily Dose; SOFAS, Social and Occupational Functioning Assessment Scale; NEET, not in employment, education and training; TLI, Thought and Language Index; MLS, mean length of sentences; MLT, mean length of T-units; MLC, mean length of clauses.

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

differences in psychotic symptoms (Oertel-Knöchel et al., 2013) and Thought and Language Disorder scores in schizophrenia (Palaniyappan et al., 2020). Prior cluster analytic studies have uncovered a consistent cluster of patients with generalized reduction in cortical thickness (Dwyer et al., 2018; Chand et al., 2020; Liang et al., 2022). We use similar methods and then we examine if these subgroups have a meaningful neurochemical basis for their differences, by examining the MRS-derived glutamate levels measured from their frontal cortex, extending our recent work (Liang et al., 2022) to a larger sample.

Abnormal cortical thickness in schizophrenia has been previously linked to dysregulated glutamate levels (Plitman et al., 2014, 2016; Shah et al., 2020) and glutamatergic dysfunction had been considered to contribute to the "FTD" burden in schizophrenia (Kircher et al., 2018). We select dACC as our region of interest for glutamate measurement as it constitutes the core hub of the large-scale brain network called the Salience Network that appears to play a key role in the neurocognitive dysfunction in schizophrenia (Palaniyappan, 2021). Finally, we used a picture description task to study computational linguistic measures that are reflective of a "negative" FTD, first described by Fish (Casey and Kelly, 2019) and later reported by Andreasen (1979) and others (Kircher et al., 2018) as being more characteristic of established schizophrenia. Negative FTD is characterized by reduced quantity and quality of speech output; in a linguistically impoverished subgroup, this will be reflected in (i) reduced fluency (number of words spoken), (ii) reduced cohesion (measured by counting instances of content with prior reference, i.e., repeat content lemmas, e.g., run, running, and ran), and (iii) reduced syntactic complexity [mean length of sentences (MLS), clauses and minimal terminable units T-units, the smallest word group that could be considered a grammatical sentence, often composed of a main clause and subordinate clauses attached to it (Hunt, 1970)].

While there are numerous quantitative linguistic measures reported to be different in case-control comparisons, we chose items that predominantly map onto the negative symptom domain of schizophrenia (Tan et al., 2021; Bilgrami et al., 2022), independent of corpus-based distributional probabilities [which has limitations in understanding compositionality (Lenci, 2018)—a crucial locus of dysfunction in schizophrenia (Chaika, 1974)] and are readily interpretable [e.g., we did not use referential cohesion measure which is conflated in the presence of perseveration (Lundin et al., 2020)]. The features we selected are also intuitive in their link to known clinical features [reduced word count relates to alogia; lack of cohesion and simplified syntax relates to the poverty of content (Bedi et al., 2015; Corcoran et al., 2018; Minor et al., 2019)]. Furthermore, compared to other aspects of communication disturbances, the features of reduced fluency and richness of content (negative factor) selectively relate to poor response to treatment (Peralta et al., 1992). A neuroanatomically defined subgroup high in

these "negative FTD type" linguistic features can be expected to be of prognostic relevance in schizophrenia.

Considering previous structural imaging-based cluster analytic studies, our primary hypothesis is that patient subgroups with distinct cortical thickness patterns can be identified in first-episode schizophrenia. In particular, a subgroup with widespread cortical thinning would emerge. Considering the association between cortical thinning, dysregulated glutamate levels and FTD burden, our secondary hypotheses are as follows: (i) The subgroup with deviant cortical thickness patterns also has abnormal glutamate levels measured in dACC; (ii) This subgroup displays impairments (negative FTD-type) in language production features, such as syntactic simplicity, reduced speech output and lower speech cohesion.

Materials and methods

Participants

We recruited 76 patients with first-episode psychosis from the Prevention and Early Intervention for Psychosis Program at the London Health Sciences Centre in London, Ontario, Canada from 2017 to 2021. Since 10 patients were unable to go through magnetic resonance imaging (MRI) scanning, we included data collected for 66 patients in this study. Inclusion criteria for patients include (1) having less than 14 days of lifetime exposure to antipsychotic medications, and (2) being at their first clinical presentation of psychotic symptoms. We followed up with patients for over 6 months to determine the validity of a diagnosis of first-episode schizophrenia prospectively. We also recruited 36 healthy volunteers, group-matched for age, sex, and parental socioeconomic status, who had no personal history of mental illnesses and no family history of psychotic disorders. All participants had no significant head injury, drug/alcohol dependence, or major medical illnesses, were fluent in English, and provided written informed consent to participate in the study. The work reported here is part of a longitudinal study registered on clinicaltrials.gov (Identifier: NCT02882204) and approved by the Western University Health Sciences Research Ethics Board, London, Ontario, Canada.

Measures and instruments

Psychiatric symptoms

Symptom severity was measured by the 8-item Positive and Negative Syndrome Scale (PANSS) (Lin et al., 2018) through interviews conducted by two research psychiatrists. Functional outcome was indexed by the Social and Occupational Functional Assessment Scale (SOFAS) (Morosini et al., 2000). The duration of untreated psychosis was calculated using the first report of positive symptoms as the starting point. We

also obtained patients' NEET (Not in Education, Employment, and Training) status. We converted participants' level of education into an ordinal scale (1: incomplete high school diploma; 2: completed high school diploma; 3: some post-secondary study; 4: completed post-secondary study or higher). Lifetime antipsychotic medication exposure was calculated by multiplying the number of days taking antipsychotics and prescribed Defined Daily Dose (DDD) values according to the World Health Organization (World Health Organization [WHO], 2022).

Thought and Language Index

Data was collected using Thought and Language Index (TLI) (Liddle et al., 2002) to reflect the two dimensions of language disorders in schizophrenia, impoverishment and disorganization. We used a picture-speech task that induced participants to elaborate 1-min spontaneous speech (oral soliloquies) in response to three images from the Thematic Apperception Test (Murray, 1943) after hearing specific instructions: "I am going to show you some pictures, one at a time. When I put each picture in front of you, I want you to describe the picture to me, as fully as you can. Tell me what you see in the picture, describe what you see in this image, and what you think might be happening." Responses were recorded, transcribed, and scored. Impoverishment score was the sum of scores for these 3 dimensions: poverty of speech, weakening of goal and preservation of ideas, while disorganization score was indexed by 5 dimensions: looseness, peculiar use of words, peculiar sentences, peculiar logic, and distractibility.

Language assessment

The same transcribed speech samples also underwent automatic analysis to measure both syntactic complexity and cohesion at the semantic level.

Tool for the automatic analysis of syntactic complexity and sophistication

The automatic analysis of syntactic complexity and sophistication (TAASSC) is an open-source[1] used in wide-ranging languages and grammatical frameworks with recent improvements in machine-learning approaches and NLP. This tool is complemented by a syntactic complexity analyzer (SCA)—a package with an accuracy of around 90% in part of speech (POS) tagging. The package includes a traditional and large measure of syntactic complexity following the taxonomy in Lu (2010): mean length of sentences (MLS), mean length of T-units (MLT), and mean length of clauses (MLC), word counts, and Terminal Units (T-unit) defined as the main clause with its attached subordinate clause(s) indicating speech cohesion as well as logical flow in the given information (see **Supplementary material** for more detailed descriptions).

Tool for the automatic analysis of cohesion

Tool for the automatic analysis of cohesion (TAACO 2.0)[2] (Crossley et al., 2016) is a freely available text analysis tool which incorporates a wide-ranging of global indices—over 150 classic and recently developed indices related to text cohesion—local, global, and overall text cohesion can significantly predict both text cohesion and speaking quality whether the speaking samples show greater semantic overlap incorporating automated semantic analysis (Crossley et al., 2019). TAACO includes 194 indices of cohesion in seven main categories: Type token ratio (TTR) and density, lexical overlap (sentences), lexical overlap (paragraphs), semantic overlap, connectives, givenness, and source text similarity. Of this, we focus on the givenness index as we analyze speech rather than written text. Givenness, as opposed to newness in a discourse transcript, indicates whether information occurring in a segment has already occurred in an earlier segment. Repeat content words or lemmas (e.g., nouns, verbs, adjectives, etc.) are calculated as a proportion of the total number of words spoken within each 1-min picture description.

Magnetic resonance imaging and magnetic resonance spectroscopy acquisition and processing

A total of 66 participants underwent neuroanatomy and spectroscopy scanning with an ultra-high-resolution 7-Tesla MRI scanner (8-channel transmit and 32-channel receive head-only coil) at Centre for Functional and Metabolic Mapping (CFMM), Western University, London, Canada. Structural images were obtained by a T1-weighted 0.75 mm isotropic MP2RAGE sequence with the following parameters: Repetition Time (TR) = 6,000 ms, Time to Echo (TE) = 2.83 ms, Inversion Time $(TI)_1$ = 800 ms, TI_2 = 2,700 ms, flip-angle 1 (α_1) = 4°, flip-angle 2 (α_2) = 5°, Field of View (FOV) = 350 mm × 263 mm × 350 mm, T_{acq} = 9 min 38 s, $iPAT_{PE}$ = 3 and 6/8 partial k-space, slice thickness = 0.75 mm. Freesurfer (version 6.0.0) (FreeSurfer Software Suite, 2021) was used to preprocess the obtained T1-weighted images. FreeSurfer provides automated brain image processing steps including intensity normalization, tissue segmentation and cortical parcellation (recon-all Free Surfer Wiki, 2022). Visual inspections of errors such as surface location misplacement were carried out according to the troubleshooting guide provided by FreeSurfer team (FsTutorial/TroubleshootingData, 2022). We acquired the cortical thickness values based on the Destrieux parcellation atlas (Destrieux et al., 2010). Magnetic resonance spectroscopy (MRS) signal was measured on a voxel placed in the dorsal anterior cingulate cortex (dACC; MNI coordinates:

---

[1] https://www.linguisticanalysistools.org/taassc.html

[2] https://www.linguisticanalysistools.org/taaco.html

TABLE 2 Demographic, clinical, neurobiological, and linguistic data of subgroups.

	Subgroup 1 patients	Subgroup 2 patients	Patient subgroup comparison	Subgroup 1 healthy controls
N	46	20		33
Demographics			Pearson's Chi-squared test or Welch t-tests	
Age (years)	21.37 (3.72)	26.15 (5.31)	$t(27.433) = -3.6527$, $p = 0.001081$ *	21.15 (3.08)
Female/male	10/36	2/18	X-squared (1) = 0.62274, $p = 0.43$	12/21
Education scale (1/2/3/4)	9/14/16/7	6/4/4/6	X-squared (3) = 3.7761, $p = 0.2867$	5/3/14/10
Clinical			Welch t-tests	
PANSS-8 (total)	25.76 (7.02)	23.85 (5.91)	$t(42.677) = 1.1376$, $p = 0.2616$	—
PANSS-8 positive	11.67 (3.46)	11.50 (3.64)	$t(34.519) = 0.18146$, $p = 0.8571$	—
PANSS-8 negative	7.48 (4.46)	5.80 (4.15)	$t(38.757) = 1.4755$, $p = 0.1481$	—
PANSS-8 general	5.22 (2.41)	5.10 (2.63)	$t(33.503) = 0.17063$, $p = 0.8655$	
DUP (weeks) [median (IQR)]	13 (4, 26)	8.5 (5.75, 16.5)	$t(23.362) = -0.53167$, $p = 0.6027$	—
DDD lifetime exposure [median (IQR)]	0 (0, 2.54)	1.25 (0, 3.9)	$t(20.156) = -1.6477$, $p = 0.1149$	—
Functional			Pearson's Chi-squared test or Welch t-tests	
SOFAS	40.98 (13.19)	40.90 (10.67)	$t(44.354) = 0.025424$, $p = 0.9798$	—
NEET status: yes/no	19/19	5/10	X-squared (1) = 0.62686, $p = 0.4285$	—
Neurobiological			ANOVA with age as a covariate	
Glutamate (mM)	6.57 (1.03)	7.28 (1.30)	$F(1) = 5.10$, $p = 0.028$, *Age effect: $p = 0.13$	6.50 (1.40)
Mean cortical thickness (mm)	2.50 (0.068)	2.32 (0.057)	$F(1) = 126.225$, $p < 0.000$, ***Age effect: $p = 0.12$	2.49 (0.061)
Language variables			Welch t-tests	
TLI (total)	1.28 (1.28)	1.93 (1.64)	$t(29.517) = -1.5629$, $p = 0.1287$	0.28 (0.40)
TLI impoverishment	0.48 (0.61)	0.79 (0.92)	$t(26.725) = -1.3843$, $p = 0.1777$	0.13 (0.23)
TLI disorganization	0.82 (1.14)	1.14 (1.37)	$t(30.974) = -0.92366$, $p = 0.3628$	0.16 (0.26)
Average total number of words	119.47 (35.45)	118.43 (47.46)	$t(24.954) = 0.084721$, $p = 0.9332$	141.53 (31.15)
MLS	14.58 (4.01)	13.91 (5.89)	$t(23.59) = 0.4227$, $p = 0.6763$	14.03 (2.67)
MLT	12.79 (3.09)	10.75 (2.20)	$t(43.928) = 2.9509$, $p = 0.005066$ **	12.45 (2.13)
MLC	7.90 (1.25)	7.30 (0.96)	$t(40.658) = 2.0284$, $p = 0.04911$ *	8.24 (1.21)
Repeated contents lemmas	0.240 (0.044)	0.204 (0.047)	$t(28.741) = 2.6991$, $p = 0.01152$ *	0.249 (0.034)

Values are reported as "Mean (SD)" unless specified otherwise.

IQR, Interquartile range; FEP, first episode psychosis; HC, healthy controls; PANSS, Positive and Negative Symptoms Scale; DUP, duration of untreated psychosis; DDD, Defined Daily Dose; SOFAS, Social and Occupational Functioning Assessment Scale; NEET, not in employment, education and training; TLI, Thought and Language Index; MLS, mean length of sentences; MLT, mean length of T-units; MLC, mean length of clauses.

*p < 0.05, **p < 0.01, ***p < 0.001.

1, 16, 38). The details of MRS acquisition and analysis have been previously described (see **Supplementary material**) and a subset of this sample has been reported in prior works (Jeon et al., 2021; Liang et al., 2022).

Statistical analyses

We applied agglomerative hierarchical clustering with Ward's method and Euclidean distance to 148 cortical thickness values [based on Destrieux parcellation atlas (Destrieux et al., 2010) output using FreeSurfer] of all 102 participants including 66 patients and 36 healthy controls. Agglomerative hierarchical clustering starts with calculating the distance (e.g., Euclidean distance) between all pairs of data objects and putting the most similar data objects into the same cluster. The newly formed clusters are then again grouped

with one another based on a linkage function (e.g., Ward's method), until all data objects merge into one single cluster. The optimal number of clusters was determined by the consensus votes from 16 clustering validity indices using NbClust (Charrad et al., 2014) in R (version 4.0.3). Pearson's chi-squared tests (with Yate's continuity correction) were used to compare categorical variables, while Welch t-tests were used to compare continuous variables. If the obtained subgroups showed difference(s) in confounding variables (e.g., age or gender), ANCOVA was used to show effects between subgroups while accounting for effects of the covariates. We used FreeSurfer to find (1) between-cluster differences in vertex-by-vertex cortical thickness while regressing out the effect of age using a general linear model, and to locate (2) cortical regionals that correlated with glutamatergic metabolic levels. The thickness values at each vertex were mapped to the surface of an average brain template, and the cortical

map was smoothed with a Gaussian kernel of 10 mm full width at half-maximum. We used Monte Carlo simulations with 1,000 permutations and a cluster-forming threshold of $P = 0.05$ (two-tailed) to correct for multiple comparisons as implemented in FreeSurfer.

Results

Demographic, clinical, linguistic, and neurobiological measurements are provided in **Table 1** and **Supplementary material**.

The cluster validity procedure of hierarchical clustering of 148 cortical thickness values of 66 patients with first-episode psychosis and 36 healthy controls suggested that a two-cluster solution is optimal (9/16 cluster validity indices). Results of clustering only patients are shown in **Supplementary material**. Proceeding with a two-cluster solution, around 70% of patients ($n = 46$) with first-episode psychosis were clustered with the majority of the healthy controls ($n = 33$) in Cluster 1, while the remaining 30% of patients ($n = 20$) were in Cluster 2 which only included 3 healthy individuals. Demographic, clinical, neurometabolite, and language functioning information of the three subgroups (Cluster 1 patients, Cluster 2 patients, and Cluster 1 healthy controls) is summarized in **Table 2** and **Supplementary Table 1**. Overall, compared to Cluster 1 patients, Cluster 2 patients have significantly older age, lower mean cortical thickens (non-significant age effect), higher glutamate concentration in dACC (non-significant age effect) as well as lower MLT (complexity) and repeated contents lemmas (cohesion) despite a preserved number of words within the given time frame (fluency). There is no significant difference between the two clusters in duration of untreated psychosis, lifetime exposure to antipsychotics, PANSS, and SOFAS scores.

Comparisons of cortical thickness between patients from the two subgroups (adjusted for age) are shown in **Figure 1**. After multiple testing corrections, patients in Cluster 1 had significantly lower thickness in 8 clusters (average area size = 410.44 mm^2) in the left hemisphere and right hemisphere, respectively (see **Supplementary Table 2** and **Figure 1** for details). Comparisons of cortical thickness between the patients and controls from Cluster 1 (adjusted for age and corrected for multiple comparisons) revealed no regional differences in thickness values, indicating that this subgroup of patients had a "healthy" cortical morphological pattern.

Multiple cortical regions were correlated with dACC glutamate levels in patients (**Figure 2**), but these correlations were not significant after multiple testing corrections. Correlation matrices of other variables of interest are presented in **Supplementary material**.

In summary, patients from Cluster 1 had similar neuroanatomical patterns to healthy controls, while patients from Cluster 2 were a distinct subgroup with widespread cortical thinning, higher glutamate concentration, and exhibited and reduced syntactic complexity and cohesion. This subgroup was thus impoverished in cortical structure as well as linguistic features.

Discussion

In the current study, we identified a subgroup of 30% of patients with first-episode schizophrenia who are distinguishable on the basis of their MRI-derived cortical thickness profiles—displaying a generalized reduction in thickness (referred to as "Subgroup 2") compared to the other group (70%) who have an unimpaired thickness profile similar to most healthy control subjects (referred to as "Subgroup 1"). Subgroup 2 is older in age at the time of the first presentation, has higher MRS-derived glutamate levels in the dorsal ACC and showed a pattern of linguistic impoverishment characterized by reduced fluency, syntactic simplicity, and repetitiveness. Taken together, these observations indicate a distinct subtype of schizophrenia that shows a pattern of cortical impoverishment along with linguistic impoverishment in the presence of higher prefrontal (dACC) glutamate levels at first presentation.

The emergence of a cortical impoverishment group showing a distributed reduction in cortical thickness compared to the other subgroup of patients and healthy controls is now a well-established feature of cluster analytical studies in schizophrenia. In a prior work where we studied two independent groups of patients with established schizophrenia as well as a part of the sample reported here, we observed a reliably identifiable subgroup of patients with cortical impoverishment (Liang et al., 2022), who did not differ from other patients in the cognitive or clinical severity. Similar findings also reported a "cortical impoverishment subgroup" at various illness stages (Sugihara et al., 2017; Dwyer et al., 2018; Chand et al., 2020; Pan et al., 2020), supporting the stability of this subtype. While the mechanistic processes underlying this structural deviation are still circumspect, based on the higher glutamate levels noted in this subgroup using 7T MRS from a dorsal ACC voxel, a putative link to excitotoxicity (Plitman et al., 2014) (or E/I dysfunction Limongi et al., 2020) can be drawn.

According to the NMDA hypofunction or glutamatergic dysregulation models of schizophrenia, higher glutamate transmission may relate to excitation-inhibition imbalance (Limongi et al., 2020) and if unchecked, may result in synaptic and neuronal loss (Wang and Qin, 2010). These cellular mechanisms have been hypothesized to underlie structural deficits in schizophrenia (Plitman et al., 2014). Multilevel genetic and physiological studies are needed to further pursue this observation. We now provide an important lead in this pursuit by identifying language dysfunction in this subgroup of schizophrenia. Additionally, we want to highlight the implications of dissecting neurobiological
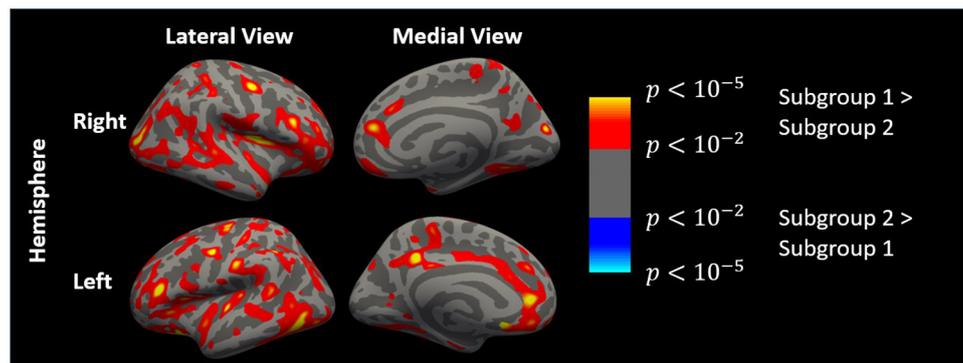
FIGURE 1
Cortical thickness map of differences between patients from Subgroup 1 and Subgroup 2 generated by FreeSurfer (regressing out age effect with a general linear model, uncorrected). Left hemisphere and right hemisphere in lateral and medial view, respectively.

heterogeneity in schizophrenia. In our study, despite displaying similar symptom severity and social functioning, the two patient subgroups have distinct neurobiological underpinnings, and may represent different pathophysiological pathways of developing schizophrenia.

Through a parts-of-speech (POS) tagging approach in NLP, we studied "poverty of content" at 3 components of grammatical structures: mean length of sentences, clauses and T-units. All are large syntactic complexity indices used as a proxy of cognitive parameters because producing a T-unit is a more complex process than producing coordinated clauses (Szmrecsanyi, 2004). T-units serve as an informative index to distinguish the amount of independent clausal coordination in the expressed idea. Moreover, T-units provided the rule-based identification process considering the selecting word for subordination (e.g., using "because") or coordination (e.g., using "and") (Beaman, 1984). Therefore, a reduction in coordinated T-units demonstrates notable syntactic simplicity in our Subgroup 2. These results are congruent with Bilgrami and colleagues' works (Bilgrami et al., 2022) who also reported lower POS syntactic complexity in those patients who had negative symptoms. The authors found that reduced sentence length and decreased use of words that introduce dependent clauses (e.g., using complementizer or determiner pronouns such as "that" and "which") are associated with negative thought disorder (Bilgrami et al., 2022). Additionally, our observations raise the question of whether patients with higher developmental disruption form the subgroup with cortical and linguistic impoverishment since syntactic complexity is a phenomenon that develops during childhood (Givon, 2009; Frizelle et al., 2018) and reaches a plateau around the age of 20 (Nippold et al., 2014). If developmental disturbances during childhood and adolescence lie in the pathogenesis of schizophrenia and can be detected using NLP tools (*via* progressive aberrations in syntactic complexity; see Silva et al., 2022), this may provide a promising avenue for early identification.

In clinical settings, linguistic dysfunction in schizophrenia relies on a standardized rating scale (PANSS and TLI) to define speech impairment as one sign of FTD (Elvevåg et al., 2007; Iter et al., 2018). The two patient subgroups did not differ in TLI or PANSS scores even though the diagnostic group of FES differed from healthy subjects in TLI rating score as expected. This observation speaks to the ability of automated quantitative processes to parse the subtler aspects of language dysfunction, an issue that has been discussed at length in several recent works based on the NLP approach (Corcoran and Cecchi, 2020; Hitczenko et al., 2021). We observed a reduction of repeated content lemma (e.g., nouns, verbs, adjectives) in our Subgroup 2. This index traditionally characterizes the systematic relationship—explicit or implicit—between lexical items, i.e., cohesive cues, placed at the text surface (Sanders and Maat, 1976). For example, if two adjacent ideas (sentence-to-sentence, clause-to-clause) comprise the same noun (e.g., woman), the lexical repetition will explicitly help connect both ideas. However, if the first clause contains the word "bridge" and the second contains the word "iron," the connection weakens even though it is logical. Therefore, in this work, we quantify cohesion (Halliday and Hasan, 1976; Graesser et al., 2004) through a lexical approach applied to how speech has been produced, without any assumption about how it is understood by listeners or readers (i.e., lexical cohesion as distinct from semantic coherence) (Just et al., 2020).

The linguistic phenomenon of reduced content word-lemmas relating to cortical thinning can be understood in several ways (Crossley et al., 2016). Firstly, reduced repetition of content-lemmas directly negatively influences the givenness of the generated speech. Givenness refers to the distribution of the given/known information or ideas as opposed to the new/unknown information. A "cortically impoverished" patient may build ideas as small clauses with little relationship between them. Secondly, a decline in the use of repeated content lemma makes it difficult to recover the
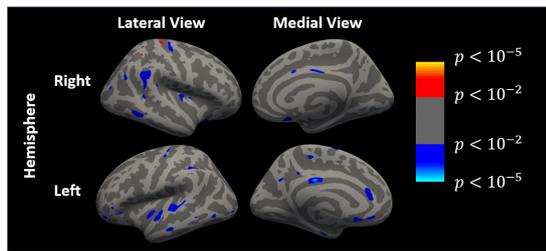
FIGURE 2

Cortical regions that are correlated with dACC glutamate levels (uncorrected) generated by FreeSurfer. Left and right hemispheres in lateral and medial view, respectively. Blue/cyan colors indicate negative correlations while red/yellow colors indicate positive correlations.

meaningful information from the preceding passage, generating a sense of empty speech (i.e., poverty of content) with reduced informative value to the listener.

Our study has several strengths: We were able to overcome the difficulty of collecting speech data in an acute, untreated state of psychosis, and determine their diagnosis of first-episode schizophrenia. Furthermore, we ensured transcribers, as well as speech analysts, were blind to diagnosis. We employed ultra-high field strength MRS whereby the glutamate quantification from MS-spectra had a high specificity. Finally, we used multiple clustering procedures and derived a two-cluster solution based on a majority-based consensus, adding to the stability of the observed subtype. Nevertheless, several limitations need consideration. We had a limited number of female participants which limits generalizability; we did not see a statistical effect of sex between the groups, but our small numbers preclude a stratified analysis. Second, thickness-based clustering resulted in age differences between the subgroups; however, we included age as a covariate in downstream analyses for glutamate and regional thickness to ensure this confound did not affect the inferences we make. Nevertheless, the non-linear influence of age on these variables cannot be ruled out. Third, we did not assess IQ formally. In our recent study where we examined the influence of cognition on thickness-based clustering in greater detail, the effect of individual differences in cognitive performance in the thickness profile was minimal among patients (Liang et al., 2022). Thus, while we can be confident that the reported thickness reduction and language dysfunction in a subgroup is not due to low extreme distributions of IQ as a result, we cannot exclude that an undetermined proportion of variance in these variables could be explained by cognitive differences. Finally, our speech samples were restricted to one language (English) and were based on a single discursive discourse (picture description) and single modality (oral soliloquies-monolog) elicited in the context of a research interview. The effect of contextual differences, language as

well as types and duration of elicitation task on our linguistic observations needs further examination.

In sum, we can link the putative excitotoxicity (glutamate excess) to reduced gray matter thickness (cortical impoverishment) and the objectively computed negative phenomenology of language (or linguistic impoverishment) in first-episode schizophrenia. This finding supports the presence of detectable neurobiological subtypes of schizophrenia. Connecting the cellular/synaptic processes (glutamate) with objectively quantified language behaviors through macroscopic brain changes (thickness) may facilitate more consistent brain-behaviors mapping in schizophrenia.

Data availability statement

The authors agree to make data and codes supporting the results presented in their paper available upon reasonable request, within the stipulations laid by the ethics committee. Requests to access the datasets should be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by the Western University Health Sciences Research Ethics Board. The patients/participants provided their written informed consent to participate in this study.

Author contributions

LP supervised and planned the study. SF and MM recruited patients and collected the data. LL and MM conducted MRI data analysis. JT designed spectroscopy protocol, and supervised PJ for MRS data acquisition and analysis. AMS analyzed the speech sample data. LL designed and conducted the statistical analyses. LL and AMS created the original draft under supervision of LP. All authors contributed to and approved the final manuscript.

Funding

Acknowledgments

Conflict of interest

LP reported personal fees from Janssen Canada, Otsuka Canada, SPMM Course Limited, United Kingdom, Canadian Psychiatric Association; book royalties from Oxford University Press; investigator-initiated educational grants from Janssen Canada, Sunovion and Otsuka Canada outside the submitted work. LP is the convenor of DISCOURSE in Psychosis, an international consortium of researchers interested in the study of language in psychosis.

The remaining 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.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum. 2022.954898/full#supplementary-material

References

Alderson-Day, B., Woods, A., Moseley, P., Common, S., Deamer, F., Dodgson, G., et al. (2021). Voice-Hearing and Personification: Characterizing Social Qualities of Auditory Verbal Hallucinations in Early Psychosis. *Schizophr. Bull.* 47, 228–236. doi: 10.1093/schbul/sbaa095

Allen, H. A., Liddle, P. F., and Frith, C. D. (1993). Negative Features, Retrieval Processes and Verbal Fluency in Schizophrenia. *Br. J. Psychiatry* 163, 769–775. doi: 10.1192/bjp.163.6.769

Alonso-Sánchez, M. F., Ford, S. D., MacKinley, M., Silva, A., Limongi, R., and Palaniyappan, L. (2022). Progressive changes in descriptive discourse in First Episode Schizophrenia: A longitudinal computational semantics study. *Schizophrenia* 8:36. doi: 10.1038/s41537-022-00246-8

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM–5)*. Virginia, US: American Psychiatric Association. doi: 10.1176/appi.books.9780890425596

Andreasen, N. C. (1979). Thought, language, and communication disorders. II. Diagnostic significance. *Arch. Gen. Psychiatry* 36, 1325–1330. doi: 10.1001/archpsyc.1979.01780120055007

Bar, K., Zilberstein, V., Ziv, I., Baram, H., Dershowitz, N., Itzikowitz, S., et al. (2019). "Semantic Characteristics of Schizophrenic Speech," in *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, (Minneapolis: Association for Computational Linguistics), 84–93. doi: 10.18653/v1/W19-3010

Beaman, K. (1984). "Coordination and subordination revisited: Syntactic complexity in spoken and written narrative discourse," in *Coherence in Spoken and Written Discourse*, ed. D. Tannen (Norwood, NJ: Ablex), 45–80.

Bedi, G., Carrillo, F., Cecchi, G. A., Slezak, D. F., Sigman, M., Mota, N. B., et al. (2015). Automated analysis of free speech predicts psychosis onset in high-risk youths. *NPJ Schizophr.* 1:15030. doi: 10.1038/npjschz.2015.30

Bilgrami, Z. R., Sarac, C., Srivastava, A., Herrera, S. N., Azis, M., Haas, S. S., et al. (2022). Construct validity for computational linguistic metrics in individuals at clinical risk for psychosis: Associations with clinical ratings. *Schizophr. Res.* 245, 90–96. doi: 10.1016/j.schres.2022.01.019

Casey, P., and Kelly, B. (2019). *Fish's Clinical Psychopathology Mental Health, Psychiatry and Clinical Psychology .4th Ed.* Cambridge: Cambridge University Press.

Chaika, E. (1974). A linguist looks at "schizophrenic" language. *Brain Lang.* 1, 257–276. doi: 10.1016/0093-934X(74)90040-6

Chaika, E., and Lambe, R. A. (1989). Cohesion in schizophrenic narratives, revisited. *J. Commun. Disord.* 22, 407–421. doi: 10.1016/0021-9924(89)90034-8

Chand, G. B., Dwyer, D. B., Erus, G., Sotiras, A., Varol, E., Srinivasan, D., et al. (2020). Two distinct neuroanatomical subtypes of schizophrenia revealed using machine learning. *Brain* 143, 1027–1038. doi: 10.1093/brain/awaa025

Charrad, M., Ghazzali, N., Boiteau, V., and Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *J. Stat. Softw.* 61, 1–36. doi: 10.18637/jss.v061.i06

Corcoran, C. M., and Cecchi, G. A. (2020). Using Language Processing and Speech Analysis for the Identification of Psychosis and Other Disorders. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 5, 770–779. doi: 10.1016/j.bpsc.2020.06.004

Corcoran, C. M., Carrillo, F., Fernández-Slezak, D., Bedi, G., Klim, C., Javitt, D. C., et al. (2018). Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry* 17, 67–75. doi: 10.1002/wps.20491

Corcoran, C. M., Mittal, V. A., Bearden, C. E., E Gur, R., Hitczenko, K., and Bilgrami, Z. (2020). Language as a biomarker for psychosis: A natural language processing approach. *Schizophr. Res.* 226, 158–166. doi: 10.1016/j.schres.2020.04.032

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

Crider, A. (1997). Perseveration in schizophrenia. *Schizophr. Bull.* 23, 63–74. doi: 10.1093/schbul/23.1.63

Crossley, S. A., Kyle, K., and Dascalu, M. (2019). The Tool for the Automatic Analysis of Cohesion 2.0: Integrating semantic similarity and text overlap. *Behav. Res. Methods* 51, 14–27. doi: 10.3758/s13428-018-1142-4

Crossley, S. A., Kyle, K., and McNamara, D. S. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behav. Res. Methods* 48, 1227–1237. doi: 10.3758/s13428-015-0651-7

de Boer, J. N., Voppel, A. E., Brederoo, S. G., Wijnen, F. N. K., and Sommer, I. E. C. (2020). Language disturbances in schizophrenia: The relation with antipsychotic medication. *NPJ Schizophr.* 6:24. doi: 10.1038/s41537-020-00114-3

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, 481–496. doi: 10.1093/oxfordjournals.schbul.a006889

Delvecchio, G., Caletti, E., Perlini, C., Siri, F. M., Andreella, A., Finos, L., et al. (2019). Altered syntactic abilities in first episode patients: An inner phenomenon characterizing psychosis. *Eur. Psychiatry* 61, 119–126. doi: 10.1016/j.eurpsy.2019.08.001

Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage* 53, 1–15. doi: 10.1016/j.neuroimage.2010.06.010

Dwyer, D. B., Cabral, C., Kambeitz-Ilankovic, L., Sanfelici, R., Kambeitz, J., Calhoun, V., et al. (2018). Brain Subtyping Enhances The Neuroanatomical Discrimination of Schizophrenia. *Schizophr. Bull.* 44, 1060–1069. doi: 10.1093/schbul/sby008

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

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

Foltz, P. W., Rosenstein, M., and Elvevag, B. (2016). Detecting clinically significant events through automated language analysis: Quo imus?. *NPJ Schizophr.* 2:15054. doi: 10.1038/npjschz.2015.54

Fraser, W. I., King, K. M., Thomas, P., and Kendell, R. E. (1986). The Diagnosis of Schizophrenia by Language Analysis. *Br. J. Psychiatry* 148, 275–278. doi: 10.1192/bjp.148.3.275

FreeSurfer Software Suite (2021). Available online at: http://surfer.nmr.mgh.harvard.edu/ (accessed May 20, 2022).

Frizelle, P., Thompson, P. A., Mcdonald, D., and Bishop, D. V. M. (2018). Growth in syntactic complexity between four years and adulthood: Evidence from a narrative task. *J. Child Lang.* 45, 1174–1197. doi: 10.1017/S0305000918000144

FsTutorial/TroubleshootingData (2022). *Free Surfer Wiki*. Available online at: https://surfer.nmr.mgh.harvard.edu/fswiki/FsTutorial/TroubleshootingData

Givon, T. (2009). *The Genesis of Syntactic Complexity*. Amsterdam: John Benjamins Publishing Company.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., and Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behav. Res. Methods Instrum. Comput.* 36, 193–202. doi: 10.3758/BF03195564

Gupta, T., Hespos, S. J., Horton, W. S., and 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. doi: 10.1016/j.schres.2017.04.025

Halliday, M. A. K., and Hasan, R. (1976). *Cohesion in English. English Language Series*. London: Longman.

Hitczenko, K., Mittal, V. A., and Goldrick, M. (2021). Understanding Language Abnormalities and Associated Clinical Markers in Psychosis: The Promise of

Computational Methods. *Schizophr. Bull.* 47, 344–362. doi: 10.1093/schbul/sbaa141

Holmlund, T. B., Fedechko, T. L., Elvevag, B., and Cohen, A. S. (2020). "Tracking language in real time in psychosis," in *A Clinical Introduction to Psychosis: Foundations for Clinical Psychologists and Neuropsychologists*, eds J. C. Badcock and G. Paulik (San Diego, CA: Elsevier Academic Press), 663–685. doi: 10.1016/B978-0-12-815012-2.00028-6

Hunt, K. W. (1970). Syntactic Maturity in Schoolchildren and Adults. *Monogr. Soc. Res. Child Dev.* 35, 1–67. doi: 10.2307/1165818

Iter, D., Yoon, J., and Jurafsky, D. (2018). "Automatic Detection of Incoherent Speech for Diagnosing Schizophrenia," in *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, (New Orleans, LA: Association for Computational Linguistics), 136–146. doi: 10.18653/v1/W18-0615

Iyegbe, C. O., and O'Reilly, P. F. (2022). Genetic origins of schizophrenia find common ground. *Nature* 604, 433–435. doi: 10.1038/d41586-022-00773-5

Jablensky, A. (2006). Subtyping schizophrenia: Implications for genetic research. *Mol. Psychiatry* 11, 815–836. doi: 10.1038/sj.mp.4001857

Jeon, P., Limongi, R., Ford, S. D., Mackinley, M., Dempster, K., Théberge, J., et al. (2021). Progressive Changes in Glutamate Concentration in Early Stages of Schizophrenia: A Longitudinal 7-Tesla MRS Study. *Schizophr. Bull. Open* 2:sgaa072. doi: 10.1093/schizbullopen/sgaa072

Just, S. A., Haegert, E., Ko¾ánová, N., Bröcker, A. L., Nenchev, I., Funcke, J., et al. (2020). Modeling Incoherent Discourse in Non-Affective Psychosis. *Front. Psychiatry* 11:846. doi: 10.3389/fpsyt.2020.00846

King, K., Fraser, W. I., Thomas, P., and Kendell, R. E. (1990). Re-examination of the Language of Psychotic Subjects. *Br. J. Psychiatry* 156, 211–215. doi: 10.1192/bjp.156.2.211

Kircher, T., Bröhl, H., Meier, F., and Engelen, J. (2018). Formal thought disorders: From phenomenology to neurobiology. *Lancet Psychiatry* 5, 515–526. doi: 10.1016/S2215-0366(18)30059-2

Kuperberg, G. R. (2010). Language in schizophrenia Part 1: An Introduction. *Lang. Linguist. Compass* 4, 576–589. doi: 10.1111/j.1749-818X.2010.00216.x

Lenci, A. (2018). Distributional Models of Word Meaning. *Annu. Rev. Linguist.* 4, 151–171. doi: 10.1146/annurev-linguistics-030514-125254

Liang, L., Heinrichs, R. W., Liddle, P. F., Jeon, P., Théberge, J., and Palaniyappan, L. (2022). Cortical impoverishment in a stable subgroup of schizophrenia: Validation across various stages of psychosis. *Schizophr. Res.* [Epub ahead of print]. doi: 10.1016/j.schres.2022.05.013

Liddle, P. F., Ngan, E. T. C., Caissie, S. L., Anderson, C. M., Bates, A. T., Quested, D. J., et al. (2002). Thought and Language Index: An instrument for assessing thought and language in schizophrenia. *Br. J. Psychiatry* 181, 326–330. doi: 10.1192/bjp.181.4.326

Limongi, R., Jeon, P., Mackinley, M., Das, T., Dempster, K., Théberge, J., et al. (2020). Glutamate and Dysconnection in the Salience Network: Neurochemical, Effective Connectivity, and Computational Evidence in Schizophrenia. *Biol. Psychiatry* 88, 273–281. doi: 10.1016/j.biopsych.2020.01.021

Lin, C. H., Lin, H. S., Lin, S. C., Kuo, C. C., Wang, F. C., and Huang, Y. H. (2018). Early improvement in PANSS-30, PANSS-8, and PANSS-6 scores predicts ultimate response and remission during acute treatment of schizophrenia. *Acta Psychiatr. Scand.* 137, 98–108. doi: 10.1111/acps.12849

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *Int. J. Corpus Linguist.* 15, 474–496. doi: 10.1075/ijcl.15.4.02lu

Lundin, N. B., Hochheiser, J., Minor, K. S., Hetrick, W. P., and Lysaker, P. H. (2020). Piecing together fragments: Linguistic cohesion mediates the relationship between executive function and metacognition in schizophrenia. *Schizophr. Res.* 215, 54–60. doi: 10.1016/j.schres.2019.11.032

Mackinley, M., Chan, J., Ke, H., Dempster, K., and Palaniyappan, L. (2021). Linguistic determinants of formal thought disorder in first episode psychosis. *Early Interv. Psychiatry* 15, 344–351. doi: 10.1111/eip.12948

Marggraf, M. P., Lysaker, P. H., Salyers, M. P., and Minor, K. S. (2020). The link between formal thought disorder and social functioning in schizophrenia: A meta-analysis. *Eur. Psychiatry* 63:e34. doi: 10.1192/j.eurpsy.2020.30

Mikesell, L., and Bromley, E. (2016). "Exploring the Heterogeneity of 'Schizophrenic Speech,'" in *The Palgrave Handbook of Adult Mental Health: Discourse and Conversation Studies*, eds M. O'Reilly and J. N. Lester (London: Palgrave Macmillan), 329–351. doi: 10.1057/9781137496850_18

Minor, K. S., Willits, J. A., Marggraf, M. P., Jones, M. N., and Lysaker, P. H. (2019). Measuring disorganized speech in schizophrenia: Automated analysis explains variance in cognitive deficits beyond clinician-rated scales. *Psychol. Med.* 49, 440–448. doi: 10.1017/S0033291718001046

Morgan, S. E., Diederen, K., Vértes, P. E., Ip, S. H. Y., Wang, B., Thompson, B., et al. (2021). Natural Language Processing markers in first episode psychosis and people at clinical high-risk. *Transl. Psychiatry* 11:630. doi: 10.1038/s41398-021-01722-y

Morice, R. D., and Ingram, J. C. (1982). Language analysis in schizophrenia: Diagnostic implications. *Aust. N. Z. J. Psychiatry* 16, 11–21. doi: 10.3109/00048678209161186

Morice, R. D., and Ingram, J. C. L. (1983). Language complexity and age of onset of schizophrenia. *Psychiatry Res.* 9, 233–242. doi: 10.1016/0165-1781(83)90048-3

Morice, R., and McNicol, D. (1986). Language Changes in Schizophrenia: A Limited Replication. *Schizophr. Bull.* 12, 239–251. doi: 10.1093/schbul/12.2.239

Morosini, P. L., Magliano, L., Brambilla, L., Ugolini, S., and Pioli, R. (2000). Development, reliability and acceptability of a new version of the DSM-IV Social and Occupational Functioning Assessment Scale (SOFAS) to assess routine social functioning. *Acta Psychiatr. Scand.* 101, 323–329. doi: 10.1111/j.1600-0447.2000.tb10933.x

Mota, N. B., Vasconcelos, N. A. P., Lemos, N., Pieretti, A. C., Kinouchi, O., Cecchi, G. A., et al. (2012). Speech Graphs Provide a Quantitative Measure of Thought Disorder in Psychosis. *PLoS One* 7:e34928. doi: 10.1371/journal.pone.0034928

Murray, H. A. (1943). *Thematic Apperception Test*. Cambridge, MA: Harvard University Press.

Nippold, M. A., Cramond, P. M., and Hayward-Mayhew, C. (2014). Spoken language production in adults: Examining age-related differences in syntactic complexity. *Clin. Linguist. Phon.* 28, 195–207. doi: 10.3109/02699206.2013.841292

Oertel-Knöchel, V., Knöchel, C., Rotarska-Jagiela, A., Reinke, B., Prvulovic, D., Haenschel, C., et al. (2013). Association between Psychotic Symptoms and Cortical Thickness Reduction across the Schizophrenia Spectrum. *Cereb. Cortex* 23, 61–70. doi: 10.1093/cercor/bhr380

Oomen, P. P., de Boer, J. N., Brederoo, S. G., Voppel, A. E., Brand, B. A., Wijnen, F. N. K., et al. (2022). Characterizing speech heterogeneity in schizophrenia-spectrum disorders. *J. Psychopathol. Clin. Sci.* 131, 172–181. doi: 10.1037/abn0000736

Oyebode, F. (2021). *Sims' Symptoms in the Mind: Textbook of Descriptive Psychopathology - 7th Edition*. Amsterdam: Elsevier.

Palaniyappan, L. (2021). Dissecting the neurobiology of linguistic disorganisation and impoverishment in schizophrenia. *Semin. Cell Dev. Biol.* doi: 10.1016/j.semcdb.2021.08.015 [Epub ahead of print].

Palaniyappan, L., Al-Radaideh, A., Gowland, P. A., and Liddle, P. F. (2020). Cortical thickness and formal thought disorder in schizophrenia: An ultra high-field network-based morphometry study. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 101:109911. doi: 10.1016/j.pnpbp.2020.109911

Palaniyappan, L., Mota, N. B., Oowise, S., Balain, V., Copelli, M., Ribeiro, S., et al. (2019). Speech structure links the neural and socio-behavioural correlates of psychotic disorders. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 88, 112–120. doi: 10.1016/j.pnpbp.2018.07.007

Pan, Y., Pu, W., Chen, X., Huang, X., Cai, Y., Tao, H., et al. (2020). Morphological Profiling of Schizophrenia: Cluster Analysis of MRI-Based Cortical Thickness Data. *Schizophr. Bull.* 46, 623–632. doi: 10.1093/schbul/sbz112

Parola, A., Lin, J. M., Simonsen, A., Bliksted, V., Zhou, Y., Wang, H., et al. (2022). Speech disturbances in schizophrenia: Assessing cross-linguistic generalizability of NLP automated measures of coherence. *medRxiv* [Preprint]. doi: 10.1101/2022.03.28.22272995

Penn, D. L., Kohlmaier, J. R., and Corrigan, P. W. (2000). Interpersonal factors contributing to the stigma of schizophrenia: Social skills, perceived attractiveness, and symptoms. *Schizophr. Res.* 45, 37–45. doi: 10.1016/S0920-9964(99)00213-3

Peralta, V., Cuesta, M. J., and de Leon, J. (1992). Formal thought disorder in schizophrenia: A factor analytic study. *Compr. Psychiatry* 33, 105–110. doi: 10.1016/0010-440X(92)90005-B

Plitman, E., Nakajima, S., de la Fuente-Sandoval, C., Gerretsen, P., Chakravarty, M. M., and Kobylianskii, J. (2014). Glutamate-mediated excitotoxicity in schizophrenia: A review. *Eur. Neuropsychopharmacol.* 24, 1591–1605. doi: 10.1016/j.euroneuro.2014.07.015

Plitman, E., Patel, R., Chung, J. K., Pipitone, J., Chavez, S., Reyes-Madrigal, F., et al. (2016). Glutamatergic Metabolites, Volume and Cortical Thickness in Antipsychotic-Naive Patients with First-Episode Psychosis: Implications for

Excitotoxicity. *Neuropsychopharmacoloy* 41, 2606–2613. doi: 10.1038/npp.2016.84

Ratana, R., Sharifzadeh, H., Krishnan, J., and Pang, S. A. (2019). Comprehensive Review of Computational Methods for Automatic Prediction of Schizophrenia With Insight Into Indigenous Populations. *Front. Psychiatry* 10:659. doi: 10.3389/fpsyt.2019.00659

recon-all Free Surfer Wiki (2022). Available online at: https://surfer.nmr.mgh.harvard.edu/fswiki/recon-all (accessed May 20, 2022).

Roche, E., Creed, L., MacMahon, D., Brennan, D., and Clarke, M. (2015). The Epidemiology and Associated Phenomenology of Formal Thought Disorder: A Systematic Review. *Schizophr. Bull.* 41, 951–962. doi: 10.1093/schbul/sbu129

Sanders, T., and Maat, H. P. (1976). *Cohesion and Coherence: Linguistic Approaches*. Available online at: https://coek.info/pdf-cohesion-and-coherence-linguistic-approaches-.html (accessed May 20, 2022).

Shah, P., Plitman, E., Iwata, Y., Kim, J., Nakajima, S., Chan, N., et al. (2020). Glutamatergic neurometabolites and cortical thickness in treatment-resistant schizophrenia: Implications for glutamate-mediated excitotoxicity. *J. Psychiatr. Res.* 124, 151–158. doi: 10.1016/j.jpsychires.2020.02.032

Silva, A. M., Limongi, R., MacKinley, M., Ford, S. D., Alonso-Sánchez, M. F., and Palaniyappan, L. (2022). Syntactic complexity of spoken language in the diagnosis of schizophrenia: A probabilistic Bayes network model. *Schizophr. Res.* [Epub ahead of print]. doi: 10.1016/j.schres.2022.06.011

Sugihara, G., Oishi, N., Son, S., Kubota, M., Takahashi, H., and Murai, T. (2017). Distinct Patterns of Cerebral Cortical Thinning in Schizophrenia: A Neuroimaging Data-Driven Approach. *Schizophr. Bull.* 43, 900–906. doi: 10.1093/schbul/sbw176

Szmrecsanyi, B. (2004). "On operationalizing syntactic complexity," in *Le poids des mots Proceedings of the 7th International Conference on Textual Data Statistical Analysis*, eds P. Gérard, F. Cédrick, and D. Anne (Louvain-la-Neuve: Presses universitaires de Louvain), 1032–1039.

Tan, E. J., Meyer, D., Neill, E., and Rossell, S. L. (2021). Investigating the diagnostic utility of speech patterns in schizophrenia and their symptom associations. *Schizophr. Res.* 238, 91–98. doi: 10.1016/j.schres.2021.10.003

Tan, E. J., Thomas, N., and Rossell, S. L. (2014). Speech disturbances and quality of life in schizophrenia: Differential impacts on functioning and life satisfaction. *Compr. Psychiatry* 55, 693–698. doi: 10.1016/j.comppsych.2013.10.016

Tang, S. X., Kriz, R., Cho, S., Park, S. J., Harowitz, J., Gur, R. E., et al. (2021). Natural language processing methods are sensitive to sub-clinical linguistic differences in schizophrenia spectrum disorders. *NPJ Schizophr.* 7:630. doi: 10.1038/s41537-021-00154-3

Thomas, P. (1996). Syntactic Complexity and Negative Symptoms in First Onset Schizophrenia. *Cogn. Neuropsychiatry* 1, 191–200. doi: 10.1080/135468096396497

Thomas, P., King, K., Fraser, W. I., and Kendell, R. E. (1990). Linguistic Performance in Schizophrenia: A Comparison of Acute and Chronic Patients. *Br. J. Psychiatry* 156, 204–210. doi: 10.1192/bjp.156.2.204

Trubetskoy, V., Pardiñas, A. F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T. B., et al. (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* 604, 502–508. doi: 10.1038/s41586-022-04434-5

Voleti, R., Liss, J. M., and Berisha, V. A. (2020). Review of Automated Speech and Language Features for Assessment of Cognitive and Thought Disorders. *IEEE J. Sel. Top. Signal. Process.* 14, 282–298.

Wang, Y., and Qin, Z. H. (2010). Molecular and cellular mechanisms of excitotoxic neuronal death. *Apoptosis* 15, 1382–1402. doi: 10.1007/s10495-010-0481-0

Wible, C. G. (2012). Schizophrenia as a Disorder of Social Communication. *Schizophr. Res. Treat.* 2012:920485. doi: 10.1155/2012/920485

World Health Organization [WHO] (2022). *Defined Daily Dose (DDD)*. Available online at: https://www.who.int/tools/atc-ddd-toolkit/about-ddd (accessed May 20, 2022).

Zhao, Y., Zhang, Q., Shah, C., Li, Q., Sweeney, J. A., Li, F., et al. (2022). Cortical Thickness Abnormalities at Different Stages of the Illness Course in Schizophrenia: A Systematic Review and Meta-analysis. *JAMA Psychiatry* 79, 560–570. doi: 10.1001/jamapsychiatry.2022.0799

Zimmerer, V. C., Watson, S., Turkington, D., Ferrier, I. N., and Hinzen, W. (2017). Deictic and Propositional Meaning—New Perspectives on Language in Schizophrenia. *Front. Psychiatry* 8:17. doi: 10.3389/fpsyt.2017.00017