REVIEW

Anomalies in language as a biomarker for schizophrenia

Janna N. de Boer[a,b], Sanne G. Brederoo[a], Alban E. Voppel[a], and Iris E.C. Sommer[a]

Purpose of review
After more than a century of neuroscience research, reproducible, clinically relevant biomarkers for schizophrenia have not yet been established. This article reviews current advances in evaluating the use of language as a diagnostic or prognostic tool in schizophrenia.

Recent findings
The development of computational linguistic tools to quantify language disturbances is rapidly gaining ground in the field of schizophrenia research. Current applications are the use of semantic space models and acoustic analyses focused on phonetic markers. These features are used in machine learning models to distinguish patients with schizophrenia from healthy controls or to predict conversion to psychosis in high-risk groups, reaching accuracy scores (generally ranging from 80 to 90%) that exceed clinical raters. Other potential applications for a language biomarker in schizophrenia are monitoring of side effects, differential diagnostics and relapse prevention.

Summary
Language disturbances are a key feature of schizophrenia. Although in its early stages, the emerging field of research focused on computational linguistics suggests an important role for language analyses in the diagnosis and prognosis of schizophrenia. Spoken language as a biomarker for schizophrenia has important advantages because it can be objectively and reproducibly quantified. Furthermore, language analyses are low-cost, time efficient and noninvasive in nature.

Keywords
language, psychosis, schizophrenia, semantic space, speech

INTRODUCTION

After more than a century of neuroscience research, reproducible, clinically relevant biomarkers for schizophrenia have not yet been established [1]. While early clinical diagnosis or relapse of a schizophrenia-spectrum disorder can be rather straightforward if there is a good working alliance between patient and psychiatrist, lack in trust, little disease insight and failing motivation may result in insufficient anamnestic information. In these situations, an objective quantitative biomarker to aid the diagnostic or prognostic process would be most welcome. However, blood-based and neuroimaging biomarkers for schizophrenia fail to reach clinically applicable levels [2–4], with diagnostic accuracies varying between 60 and 90%. A rich source of information that has so far rarely been used, is spoken language. Recent advances in the field of computational linguistics afford the clinician to turn to language output as a novel biomarker that is low-cost, time efficient and noninvasive in nature [5].

Language as a biomarker has important advantages over traditional biomarkers such as blood markers or imaging, because it can be reproducibly quantified without special training.

It has long been observed that schizophrenia is characterized by disturbed language, with Kraepelin

[a]University of Groningen, University Medical Center Groningen, department of Neuroscience and department of Psychiatry, Groningen, the Netherlands and [b]Department of Psychiatry, University Medical Center Utrecht Brain Center, Utrecht University, Utrecht, The Netherlands

Correspondence to Janna N. de Boer, MD, Department of Psychiatry, University Medical Center Utrecht, Utrecht University, Utrecht 3508 GA, The Netherlands. Tel: +31 887550163;
e-mail: j.n.deboer-15@umcutrecht.nl

KEY POINTS

- Language disturbances are a key feature of schizophrenia.

- Current advances in computational linguistic allow for the development of fast, objective tools to aid in early and correct diagnosis of schizophrenia-spectrum disorders.

- Analysing the meaning and coherence of language, as well as measuring nonverbal acoustic aspects, are promising angles towards the development of such a tool.

- These features are used to distinguish patients with schizophrenia from healthy controls or to predict conversion to psychosis in high-risk groups, reaching accuracy scores that exceed clinical raters (generally ranging from 80 to 90%).

- By combining different linguistic techniques these tools might be used for early recognition, treatment response or relapse prediction in clinical practice.

describing a subgroup of patients with 'schizophasia' [6], and Bleuler who stressed the importance of aberrant language as a feature of schizophrenia [7]. Pioneers in this line of research applied manual linguistic analyses to spoken language to evaluate its use in the diagnostic or prognostic process in schizophrenia-spectrum disorders [8–10].

Here, we reviewed the use of computational language analysis in schizophrenia-spectrum disorders with an emphasis on how recent translational research contributes to the development of diagnostic and prognostic tools. Much of the recent literature relates to advances in methodological and analytic tools which may facilitate diagnosis and prognosis of schizophrenia-spectrum disorders.

LANGUAGE DISTURBANCES IN SCHIZOPHRENIA

Impaired verbal communication is one of the key diagnostic features of schizophrenia-spectrum disorders. For reviews on this topic refer to [11–16]. Overall, patients with schizophrenia display a broad range of semantic (i.e. meaning in language) processing disturbances; including difficulties with lexical selection and retrieval [17], disturbances in priming [18] and reduced proactive inhibition [12,19]. On a discourse level, they show difficulties with coherently generating a narrative, which is thought to reflect an underlying disturbance in taking viewpoints or perspectives [20■]. Other related disturbances in schizophrenic language include: neologisms, word approximation [12], disturbances in cohesion [21],

vague references, missing information and confused references [15,22]. At a syntactic (i.e. grammatical) level, patients with a schizophrenia-spectrum disorder produce sentences with reduced syntactic complexity [23], less dependents and embedded clauses [24■], and use fewer connective markers [25]. Furthermore, syntactic priming appears to be reduced [26]. Spontaneous abnormal morphology (i.e. using abnormal word forms) in schizophrenia is quite rare [11]. In a test setting, however, patients make more morphological errors than controls [27]. Schizophrenic speech usually has normal segmental phonology (i.e. the articulation of segments such as syllables), although compared with normal speech it contains more hesitations and pauses and longer pauses [28,29], and the intonation is flat (monotonous) [12].

It has been suggested that language disturbances in schizophrenia arise from abnormal semantic and phonological processing [30–33]. Indeed, neuroimaging data implicate altered frontotemporal semantic and phonological networks in schizophrenia. These include abnormalities in the structure of Broca's, Wernicke's and other frontotemporal regions [34,35], abnormal white matter language tracts [36–40] and altered functional MRI activation patterns in a variety of language tasks [41–44]. White matter language tract alterations were found in individuals at clinical high-risk (CHR) for psychosis [45,46], suggesting that these abnormalities precede schizophrenia onset. Indeed, retrospective studies suggest childhood language delays in people who later developed schizophrenia [47,48]. Previous reports have indicated that genetic alternations underlie the neurodevelopment of language abnormalities in schizophrenia [49,50]. The first identified gene involved in language was the *FOXP2* gene [51]. Preliminary association studies on FOXP2 polymorphisms and schizophrenia have delivered inconsistent results [52–54], although epigenetic data do suggest that FOXP2 may be involved in language disorders in schizophrenia [55]. Furthermore, variations in another gene, dysbindin 1 (*DTNBP1*) have been associated with neural correlates of language production [56]. However, further research is needed to confirm these preliminary results.

Summarizing, biological correlates of language disturbances in schizophrenia have been found in both neuroimaging and genetic studies. Previous research into aberrant language in schizophrenia-spectrum disorders has investigated difficulties arising at the semantic, syntactic and phonological levels of language production. Correspondingly, computational language analyses have focused on these aspects of language output (i.e. semantics, syntax and phonology).

COMPUTATIONAL LANGUAGE ANALYSIS IN SCHIZOPHRENIA

Content analyses: meaning, structure and coherence

An often used method to examine meaning and coherence in language is that of semantic space models. Semantic space models, of which latent semantic analysis (LSA) [57] is the most commonly used tool, aim to capture word meaning by representing words as so called 'vectors' in a 'semantic space'. These vectors contain word features (i.e. aspects of word meaning); 'furry', 'pet' and 'purring' might be features attempting to grasp the meaning of 'cat'. The distance between words in a semantic space indicates word interrelatedness or coherence; the word 'furry' will be more closely related to 'pet' than to 'banana', by virtue of what concepts these words are taken to represent. A sentence with low internal coherence will consist of words reflecting relatively more separated concepts. So-called distributed models like Word2vec aim at capturing both semantic as well as syntactic information [58,59].

Spoken language

The first to introduce semantic space models in schizophrenia were Elvevåg *et al.* [60], who used LSA to show that schizophrenia patients could successfully be distinguished from healthy controls based solely on their spoken language output (achieving correct classification of patients and controls with an accuracy of 82.4%). Furthermore, this study showed that patients with formal thought disorder (FTD) could be distinguished from patients with low FTD scores (with an accuracy of 87.5%). LSA thus appears to be an accurate tool for detecting FTD. Significantly, clinical raters achieved slightly lower classification scores (84%) than the LSA models. This research was later expanded on by classifying patients with schizophrenia and their healthy family members [61]. Using cross-validation, 85.7% of patients with schizophrenia could be correctly distinguished from their family members, indicating that LSA is sensitive to subtle phenomena, as patients are taken to resemble family members more than nonfamily controls.

In their seminal study, Bedi *et al.* [62] used LSA and two measures of language complexity [maximum phrase length and the use of determiners (e.g. *that*)] on spoken language samples, to predict later psychosis onset in youths at CHR for psychosis. Combined, these language measures predicted psychosis development with 100% accuracy, outperforming clinical ratings (yielding an accuracy of 79%). However, in their sample of 34 CHR youths,

only five transitioned to psychosis. This model was adapted and validated in a larger sample, and across cohorts in a larger sample [63▪▪]. Using decreased semantic coherence, greater variance in coherence and reduced use of possessive pronouns; 83% accuracy was achieved within the main cohort (79% across cohorts).

Using a pretrained set of vectors (fastText [64], Bar *et al.* [65▪]) examined patients with schizophrenia and controls with a special emphasis on their use of adjectives and adverbs. Their results show that patients with schizophrenia use adjectives and adverbs that are less common (i.e. lower frequency words), which can be used to distinguish them from healthy controls with machine learning models (accuracies depending on the model ranging from 70.4 to 81.5%).

In a recent meta-analysis of the diagnostic and prognostic value of semantic space models [66], a large effect size was found for diagnosing schizophrenia-spectrum disorders using semantic space (Hedges' $g = 0.96$, $P = 0.003$). Semantic space models perform better on (semi) spontaneous language or sentences, than they do on lists of single words (e.g. words produced during a verbal fluency task). Pooling all studies in a meta-analysis of diagnostic test accuracy in schizophrenia-spectrum patients, an overall sensitivity of 71% and specificity of 91% was found.

Another influential approach to model coherence in language is the use of speech graphs [67–69]. Using graph-based tools to visualize connectedness in language, patients with schizophrenia could be distinguished from manic patients with a sensitivity and specificity of 94% [69].

Written language

Posts on social media have been analysed to examine written language in schizophrenia-spectrum disorders in several studies. Using content on the social media platform Reddit, conversion to psychosis was shown to be signalled by low semantic density, a measure developed to quantify sentence richness (calculated using Word2vec). Combined with writing about voices and sounds, these variables predicted conversion to psychosis with 93% accuracy [70▪].

In a similar study, Twitter content of self-proclaimed schizophrenia patients was analysed using the semantic space model Latent Dirichlet Allocation [71], in addition to part-of-speech, pragmatic analyses and syntactic dependency measures [72]. Combined, these measures were used to classify schizophrenia patients and matched controls using machine learning (support vector machine), which

resulted in an area under the curve of 82.6; indicating 83% of cases could be successfully distinguished from controls.

Further, Facebook content and behaviour analysis of patients with recent onset psychosis was used to predict relapse hospitalization [73■]. The increased use of first and second-person pronouns, swear words and words related to anger and death, as well as decreased use of words related to work, friends and health, were predictive of relapse. Combined with other behaviour on Facebook, relapse could be predicted with 71% specificity, however, sensitivity was low (38%).

Nonverbal and phonetic analyses

Computerized analyses of phonetic features (i.e. speech sounds) have also been used to objectively evaluate (especially negative) symptoms in schizophrenia-spectrum disorders. For instance, schizophrenia patients with clinically rated aprosody were shown to differ from controls in pitch variation [74]. Nonverbal language measures (e.g. turn duration, percentage of time speaking) were used to classify patients with schizophrenia and healthy controls, with an accuracy of 81.3% [75■■]. A similar study [76] measured prosodic and phonetic cues (prosodic peaks, syllabic dynamics) while reading the first paragraph of 'Don Quixote' to classify patients with schizophrenia and controls, reaching a sensitivity of 95.6% and a specificity of 91.4%, with an overall accuracy of 93.8%.

FOOD FOR THOUGHT

Language versus speech

Two important and notably distinct concepts in this line of research are 'language' and 'speech'. Language is the term used for the mental system underlying verbal behaviour, which includes meaning, grammar and form. Speech is the term used for the spoken output or the *medium* of the language, the way it is produced by the speech organs. Language can of course also be produced in writing or in gestures (sign language), which still requires similar cognitive processes to formulate sentences, without the use of the vocal tract (i.e. without articulation). Although communication difficulties in schizophrenia are currently described as 'disorganized speech', the literature discussed in this review clearly demonstrates that patients with schizophrenia display a wide variety of language disorders including broad disturbances in semantics, pragmatics and grammatical structures [12,15]. 'Disorganized speech' [77] would therefore, better be described as 'disturbed language', which may include, but is not limited to, speech.

Biomarker

The term biomarker is classically used for analytes of a human biological system (e.g. plasma, urine, cerebrospinal fluid) or for biological properties (i.e. mass concentration). However, the Biomarkers Definitions Working Group and other initiatives have advocated a broader, less ambiguous, definition of biomarkers, namely: 'a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacological response' [78,79]. Language output fully adheres to this definition and can thus serve as a true biomarker for schizophrenia-spectrum disorders.

Current state of research

Of note, in most classification models discussed in this review, the final model included one or several variables which are nonspecific to language. Examples of such general features are task duration (reading 400 words aloud) [76] and response time to a question [75■■]. These variables are most likely based on general cognitive deficits such as reductions in attention, working memory or general fatigue, which are common in schizophrenia [80]. The decision to add less specific measures to a model is presumably motivated by the aspiration of models with high diagnostic or prognostic accuracy and the pursuit of developing clinically valuable tools. However, whereas general cognitive measures may have high discriminatory power, employing them in an early stage forecloses improvement of our knowledge of language-related disturbances in schizophrenia. Further, including nonspecific measures in classification models reduces their power to detect early or subtle symptoms in spoken language that are specific to schizophrenia and may be used for differential diagnosis. While we endorse the ultimate goal of developing highly accurate diagnostic and prognostic tools, the aim to assess the value of purely linguistic measures should not be neglected. To this end, results of models with only linguistic features should be reported as well, even if they are less accurate than models that in addition include nonspecific factors.

A related point of discussion is that in extensive machine learning and deep learning models, features become abstract and an abounding number of features is fed to the model (e.g. 40 526 speech features were used in a model to detect post-traumatic stress disorder [81]), which renders it difficult

to retrace a classification model to clinically recognizable symptoms or signs. A word of caution for this development is, therefore, in order. In an extreme example such tendencies could lead to a model that bases its classification of patients and healthy controls only on their use of antipsychotic medication. This of course would lead to (near) perfect classification scores, but such a model would have no diagnostic value. Similarly, algorithms might 'overfit' predictions due to for example multicollinearity or correlated predictors, producing unstable estimates. Such problems can be overcome by validation in a truly independent dataset; problems in the model fitting stage will show up as poor performance in a validation process. However, of the studies reviewed here, most use cross-validation to assess the generalizability of their models, which does not fully overcome this risk of overfitting. Few studies validated their models in a separate subset of their data [70▪] or in an independent dataset [63▪▪]; the latter of which should become the standard in this field of research.

CONCLUSION

The value of computational language analyses as biomarkers in schizophrenia-spectrum disorders is increasing as a result of rapidly advancing linguistic techniques. Language technology evolves quickly and analytic techniques such as machine learning allow for the application of complex features to a clinically relevant goal. Language analyses show potential for a range of applications in schizophrenia; for example in identifying at risk groups on social media [82,83], monitoring psychosis relapse through smartphone applications [84] or predicting treatment response. Recent work using computational semantic tools such as semantic space and graph analysis, as well as phonetic acoustic markers, have proved successful in both diagnosis and prognosis of schizophrenia-spectrum disorders. Accuracy scores in differentiating patients from healthy controls, family members or at risk groups range from 80 to 90%, often outperforming clinical raters. Even the clinically difficult differentiation between psychosis and mania showed high specificity and sensitivity with language analysis (both 94%).

Further longitudinal studies across a broader range of ages, disease severity and illness durations will be needed to understand the trajectory of language disturbances in schizophrenia-spectrum disorders. Future research is needed to fully appraise the potential of language as a diagnostic or prognostic tool. For example, a variety of language characteristics could be targeted by combining disparate computational tools. This may improve the predictive power substantially; since the most often used tools (semantic space and acoustic measures) are thought to be a reflection of a different set of symptoms. Semantic incoherence is often associated with FTD or disorganized language [24▪,60,85], while acoustic measures are often used to objectify negative symptoms [29,75▪▪,86,87]. Bringing these methods together acknowledges the heterogeneity of symptoms associated with schizophrenia-spectrum disorders. Combining several quantifiable aspects of language may also pave the road towards cross-diagnostic analyses. Finally, researchers in this field should aim to do cross-linguistic analyses, to examine whether these models hold for the great diversity of languages in the world.

Conflicts of interest

I.S. is a consultant to Gabather, received research support from Janssen Pharmaceuticals Inc. and Sunovion Pharmaceuticals Inc.

REFERENCES AND RECOMMENDED READING
Papers of particular interest, published within the annual period of review, have been highlighted as:
■ of special interest
■■ of outstanding interest

1. Insel TR. The NIMH research domain criteria (RDoC) project: precision medicine for psychiatry. Am J Psychiatry 2014; 171:395–397.
2. Kambeitz J, Kambeitz-Ilankovic L, Leucht S, et al. Detecting neuroimaging biomarkers for schizophrenia: a meta-analysis of multivariate pattern recognition studies. Neuropsychopharmacology 2015; 40:1742.
3. Zarogianni E, Moorhead TWJ, Lawrie SM. Towards the identification of imaging biomarkers in schizophrenia, using multivariate pattern classification at a single-subject level. NeuroImage Clin 2013; 3:279–289.
4. Chan MK, Krebs M-O, Cox D, et al. Development of a blood-based molecular biomarker test for identification of schizophrenia before disease onset. Transl Psychiatry 2015; 5:e601.
5. Tan E, Rossell S. Questioning the status of aberrant speech patterns as psychiatric symptoms. Br J Psychiatry. 2020; In press.
6. Kraepelin E. Dementia praecox and paraphrenia (RM Barclay, Trans). Huntingdon, NY: Krieger; 1919.
7. Bleuler E. Dementia praecox oder Gruppe der Schizophrenien. 1st ed. Leipzig, Wien: Diskord; 1911.
8. Fraser WI, King KM, Thomas P, Kendell RE. The diagnosis of schizophrenia by language analysis. Br J Psychiatry 1986; 148:275–278.
9. King K, Fraser WI, Thomas P, Kendell RE. Re-examination of the language of psychotic subjects. Br J Psychiatry 1990; 156:211–215.
10. Morice RD, Ingram JCL. Language analysis in schizophrenia: diagnostic implications. Aust N Z J Psychiatry 1982; 16:11–21.
11. Chaika E. Understanding psychotic speech: beyong Freud and Chomsky. Springfield, Illinois: Charles C. Thomas; 1990.
12. Covington MA, He C, Brown C, et al. Schizophrenia and the structure of language: the linguist's view. Schizophr Res 2005; 77:85–98.

13. DeLisi LE. Speech disorder in schizophrenia: review of the literature and exploration of its relation to the uniquely human capacity for language. Schizophr Bull 2001; 27:481–496.
14. Docherty NM, DeRosa M, Andreasen NC. Communication disturbances in schizophrenia and mania. Arch Gen Psychiatry 1996; 53:358–364.
15. Kuperberg GR. Language in schizophrenia Part 1: An introduction. Lang Linguist Compass 2010; 4:576–589.
16. Kuperberg GR. Language in schizophrenia Part 2. Lang Linguist Compass 2011; 4:576–589.
17. Tan EJ, Neill E, Rossell SL. Assessing the relationship between semantic processing and thought disorder symptoms in schizophrenia. J Int Neuropsychol Soc 2015; 21:629–638.
18. Kuperberg GR, Delaney-Busch N, Fanucci K, Blackford T. Priming production: neural evidence for enhanced automatic semantic activity preceding language production in schizophrenia. NeuroImage Clin 2018; 18:74–85.
19. Maher BA, Manschreck TC, Linnet J, Candela S. Quantitative assessment of the frequency of normal associations in the utterances of schizophrenia patients and healthy controls. Schizophr Res 2005; 78:219–224.
20. ■ van Schuppen L, van Krieken K, Sanders J. Deictic navigation network: linguistic viewpoint disturbances in schizophrenia. Front Psychol 2019; 10:1616.
Interesting study that proposes a cognitive-linguistic framework to understand viewpoint disturbances in schizophrenia.
21. Docherty NM, DeRosa M, Andreasen NC. Commuication disturbances in schizophrenia and mania. Arch Gen Psychiatry 1996; 53:358–364.
22. Ditman T, Kuperberg GR. Building coherence: a framework for exploring the breakdown of links across clause boundaries in schizophrenia. J Neurolinguistics 2010; 23:254–269.
23. Özcan A, Kuruoglu G, Alptekin K, et al. The production of simple sentence structures in schizophrenia. Int J Arts Sci 2017; 09:159–164.
24. ■ Çokal D, Sevilla G, Jones WS, et al. The language profile of formal thought disorder. NPJ Schizophr 2018; 4:18.
Important study that shows that formal thought disorder can be identified using computational linguistic tools.
25. Willits JA, Rubin T, Jones MN, et al. Evidence of disturbances of deep levels of semantic cohesion within personal narratives in schizophrenia. Schizophr Res 2018; 197:365–369.
26. Dwyer K, David AS, McCarthy R, et al. Linguistic alignment and theory of mind impairments in schizophrenia patients' dialogic interactions. Psychol Med 2019; 1–9. [Epub ahead of print]
27. Walenski M, Weickert TW, Maloof CJ, Ullman MT. Grammatical processing in schizophrenia: evidence from morphology. Neuropsychologia 2010; 48:262–269.
28. Clemmer EJ. Psycholinguistic aspects of pauses and temporal patterns in schizophrenic speech. J Psycholinguist Res 1980; 9:161.
29. Cohen AS, Mitchell KR, Elvevåg B. What do we really know about blunted vocal affect and alogia? A meta-analysis of objective assessments. Schizophr Res 2014; 159:533–538.
30. Kuperberg GR, Goldberg T. Insights into semantics and language in schizophrenia. In: Handbook of Neuropsychology of Mental Illness. Cambridge, UK: Cambridge University Press; 2006.
31. Kreher DA, Holcomb PJ, Goff D, Kuperberg GR. Neural evidence for faster and further automatic spreading activation in schizophrenic thought disorder. Schizophr Bull 2007; 34:473–482.
32. Angrilli A, Spironelli C, Elbert T, et al. Schizophrenia as failure of left hemispheric dominance for the phonological component of language. PLoS One 2009; 4:e4507.
33. Spitzer M, Weisker I, Winter M, et al. Semantic and phonological priming in schizophrenia. J Abnorm Psychol 1994; 103:485–494.
34. Sans-Sansa B, McKenna PJ, Canales-Rodríguez EJ, et al. Association of formal thought disorder in schizophrenia with structural brain abnormalities in language-related cortical regions. Schizophr Res 2013; 146:308–313.
35. Wisco JJ, Kuperberg G, Manoach D, et al. Abnormal cortical folding patterns within Broca's area in schizophrenia: evidence from structural MRI. Schizophr Res 2007; 94:317–327.
36. Shenton ME, Dickey CC, Frumin M, McCarley RW. A review of MRI findings in schizophrenia. Schizophr Res 2001; 49:1–52.
37. Catani M, Craig MC, Forkel SJ, et al. Altered integrity of perisylvian language pathways in schizophrenia: relationship to auditory hallucinations. Biol Psychiatry 2011; 70:1143–1150.
38. Kubicki M, Alvarado JL, Westin C-F, et al. Stochastic tractography study of inferior frontal gyrus anatomical connectivity in schizophrenia. Neuroimage 2011; 55:1657–1664.
39. Cavelti M, Kircher T, Nagels A, et al. Is formal thought disorder in schizophrenia related to structural and functional aberrations in the language network? A systematic review of neuroimaging findings. Schizophr Res 2018; 199:2–16.
40. Cavelti M, Winkelbeiner S, Federspiel A, et al. Formal thought disorder is related to aberrations in language-related white matter tracts in patients with schizophrenia. Psychiatry Res Neuroimaging 2018; 270:40–50.
41. Sommer IEC, Ramsey NF, Mandl RCW, Kahn RS. Language lateralization in female patients with schizophrenia: an fMRI study. Schizophr Res 2003; 60:183–190.
42. Sommer IEC, Ramsey NF, Kahn RS. Language lateralization in schizophrenia, an fMRI study. Schizophr Res 2001; 52:57–67.
43. Li X, Branch CA, DeLisi LE. Language pathway abnormalities in schizophrenia: a review of fMRI and other imaging studies. Curr Opin Psychiatry 2009; 22:131–139.
44. Kuperberg GR, West WC, Lakshmanan BM, Goff D. Functional magnetic resonance imaging reveals neuroanatomical dissociations during semantic integration in schizophrenia. Biol Psychiatry 2008; 64:407–418.
45. Li X, Wu K, Zhang Y, et al. Altered topological characteristics of morphological brain network relate to language impairment in high genetic risk subjects and schizophrenia patients. Schizophr Res 2019; 208:338–343.
46. Thermenos HW, Whitfield-Gabrieli S, Seidman LJ, et al. Altered language network activity in young people at familial high-risk for schizophrenia. Schizophr Res 2013; 151:229–237.
47. Hoff AL, Svetina C, Shields G, et al. Ten year longitudinal study of neuropsychological functioning subsequent to a first episode of schizophrenia. Schizophr Res 2005; 78:27–34.
48. DeLisi LE, Boccio AM, Riordan H, et al. Familial thyroid disease and delayed language development in first admission patients with schizophrenia. Psychiatry Res 1991; 38:39–50.
49. Crow TJ. Schizophrenia as the price that homo sapiens pays for language: a resolution of the central paradox in the origin of the species. Brain Res Brain Res Rev 2000; 31:118–129.
50. Crespi B, Summers K, Dorus S. Adaptive evolution of genes underlying schizophrenia. Proc R Soc B Biol Sci 2007; 274:2801–2810.
51. Lai CSL, Fisher SE, Hurst JA, et al. A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 2001; 413:519.
52. Sanjuan J, Tolosa A, González JC, et al. FOXP2 polymorphisms in patients with schizophrenia. Schizophr Res 2005; 73:253–256.
53. Sanjuán J, Tolosa A, González JC, et al. Association between FOXP2 polymorphisms and schizophrenia with auditory hallucinations. Psychiatr Genet 2006; 16:67–72.
54. Jung SM, Jung BJ, Cho JS, Park JM. P. 3. a. 007 FOXP2 gene possibly associated with Korean schizophrenic patients. Eur Neuropsychopharmacol 2008; 18:S389.
55. Tolosa A, Sanjuan J, Dagnall AM, et al. FOXP2 gene and language impairment in schizophrenia: association and epigenetic studies. BMC Med Genet 2010; 11:114.
56. Markov V, Krug A, Krach S, et al. Genetic variation in schizophrenia-risk-gene dysbindin 1 modulates brain activation in anterior cingulate cortex and right temporal gyrus during language production in healthy individuals. Neuroimage 2009; 47:2016–2022.
57. Landauer T, Dumais S. A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychol Rev 1997; 104:211–240.
58. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv 2013; 1301.3781:1–12.
59. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems. 2013. p. 3111–9.
60. Elvevåg B, Foltz PW, Weinberger DR, Goldberg TE. Quantifying incoherence in speech: an automated methodology and novel application to schizophrenia. Schizophr Res 2007; 93:304–316.
61. Elvevåg B, Foltz PW, Rosenstein M, DeLisi LE. An automated method to analyze language use in patients with schizophrenia and their first-degree relatives. J Neurolinguistics 2010; 23:270–284.
62. Bedi G, Carrillo F, Cecchi GA, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophr 2015; 1: 15030.
63. ■■ Corcoran CM, Carrillo F, Fern D, et al. Prediction of psychosis across protocols and risk cohorts using automated language analysis. World Psychiatry 2018; 17:67–75.
A pivotal study that demonstrates the use of semantic space models as a prognostic tool for psychosis development in at risk-groups and validates their algorithm in an independent sample.
64. Grave E, Bojanowski P, Gupta P, Joulin A, Mikolov T. Learning word vectors for 157 languages. Lr 2018 – 11th Int Conf Lang Resour Eval. 2019; 3483–7.
65. ■ Bar K, Zilberstein V, Ziv I, et al. Semantic characteristics of schizophrenic speech. arXiv preprint arXiv:1904.07953.2019; 84–93.
The study demonstrates that that patients with schizophrenia use adjectives and less common adverbs, which distinguishes them from healthy controls.
66. de Boer JN, Voppel AE, Begemann MJH, et al. Clinical use of semantic space models in psychiatry and neurology: a systematic review and meta-analysis. Neurosci Biobehav Rev 2018; 93:85–92.
67. Palaniyappan L, Mota NB, Oowise S, et al. Speech structure links the neural and socio-behavioural correlates of psychotic disorders. Prog Neuropsychopharmacol Biol Psychiatry 2019; 88:112–120.
68. Mota NB, Furtado R, Maia PPC, et al. Graph analysis of dream reports is especially informative about psychosis. Sci Rep 2014; 4:1–7.
69. Mota NB, Vasconcelos NAP, Lemos N, et al. Speech graphs provide a quantitative measure of thought disorder in psychosis. PLoS One 2012; 7:1–9.

70. Rezaii N, Walker E, Wolff P. A machine learning approach to predicting
■ psychosis using semantic density and latent content analysis. NPJ Schizophr
2019; 5:9.
An important study that shows that the Word2vec applied to content on social
media platforms can predict conversion to psychosis.

71. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res
2003; 3:993–1022.

72. Kayi ES, Diab M, Pauselli L, Compton M, Coppersmith G. Predictive linguistic
features of schizophrenia. *SEM 2017 - 6th Jt Conf Lex Comput Semant Proc
2017;241–50

73. Birnbaum ML. Detecting relapse in youth with psychotic disorders utilizing
■ patient-generated and patient-contributed digital data from Facebook. NPJ
Schizophr 2019; 5:17.
The study shows that content analysis of Facebook posts can be used to predict
relapse.

74. Compton MT, Lunden A, Cleary SD, et al. The aprosody of schizophrenia:
computationally derived acoustic phonetic underpinnings of monotone
speech. Schizophr Res 2018; 197:392–399.

75. Tahir Y, Yang Z, Chakraborty D, et al. Nonverbal speech cues as objective
■■ measures for negative symptoms in patients with schizophrenia. PLoS One
2019; 14:1–17.
Important study that shows that nonverbal language measures can be used to
accurately classify patients with schizophrenia and healthy controls.

76. Martínez-Sánchez F, Muela-Martínez JA, Cortés-Soto P, et al. Can the
acoustic analysis of expressive prosody discriminate schizophrenia? Span
J Psychol 2015; 18:E86.

77. American Psychiatric Association. Diagnostic and statistical manual of mental
disorders (DSM-5®). Arlington, VA: American Psychiatric Pub; 2013.

78. Biomarkers Definitions Working Group. Biomarkers and surrogate end
points: preferred definitions and conceptual framework. Clin Pharmacol Ther
2001; 69:89–95.

79. Weickert CS, Weickert TW, Pillai A, Buckley PF. Biomarkers in schizophrenia:
a brief conceptual consideration. Dis Markers 2013; 35:3–9.

80. Heinrichs RW, Zakzanis KK. Neurocognitive deficit in schizophrenia: a quan-
titative review of the evidence. Neuropsychology 1998; 12:426.

81. Marmar CR, Brown AD, Qian M, et al. Speech-based markers for posttrau-
matic stress disorder in US veterans. Depress Anxiety 2019; 36:607–616.

82. Coppersmith G, Leary R, Crutchley P, Fine A. Natural language processing of
social media as screening for suicide risk. Biomed Inform Insights 2018;
10:1178222618792860.

83. Robinson J, Cox G, Bailey E, et al. Social media and suicide prevention: a
systematic review. Early Interv Psychiatry 2016; 10:103–121.

84. Holmlund T, Foltz PW, Cohen AS, et al. Moving speech technology methods
out of the laboratory: practical challenges and clinical translation opportunities
for psychiatry. Schizophr Bull 2019; 45(Suppl_2):S129–S1129.

85. Bedi G, Cecchi GA, Slezak DF, et al. A window into the intoxicated mind?
Speech as an index of psychoactive drug effects. Neuropsychopharmacology
2014; 39:2340–2348.

86. Covington MA, Lunden SLA, Cristofaro SL, et al. Phonetic measures of
reduced tongue movement correlate with negative symptom severity in
hospitalized patients with first-episode schizophrenia-spectrum disorders.
Schizophr Res 2012; 142:93–95.

87. Bernardini F, Lunden A, Covington M, et al. Associations of acoustically
measured tongue/jaw movements and portion of time speaking with negative
symptom severity in patients with schizophrenia in Italy and the United States.
Psychiatry Res 2016; 239:253–258.