

Latent Factors of Language Disturbance and Relationships to Quantitative Speech Features

Sunny X. Tang^{*,1,✉}, Katrin Hänsel², Yan Cong¹, Amir H. Nikzad¹, Aarush Mehta¹, Sunghye Cho³, Sarah Berretta¹, Leily Behbehani¹, Sameer Pradhan³, Majnu John¹, and Mark Y. Liberman³

¹Institute of Behavioral Science, Feinstein Institutes for Medical Research, Glen Oaks, USA²Department of Laboratory Medicine, Yale University, New Haven, USA³Linguistic Data Consortium, University of Pennsylvania, Philadelphia, USA

*To whom correspondence should be addressed; Zucker Hillside Hospital; Psychiatry Research Division; 75–59 263rd St.; Glen Oaks, NY 11004, USA; tel: 718-470-8267, fax: 7183431659, e-mail: stang3@northwell.edu

Background and Hypothesis: Quantitative acoustic and textual measures derived from speech (“speech features”) may provide valuable biomarkers for psychiatric disorders, particularly schizophrenia spectrum disorders (SSD). We sought to identify cross-diagnostic latent factors for speech disturbance with relevance for SSD and computational modeling. **Study Design:** Clinical ratings for speech disturbance were generated across 14 items for a cross-diagnostic sample ($N = 343$), including SSD ($n = 90$). Speech features were quantified using an automated pipeline for brief recorded samples of free speech. Factor models for the clinical ratings were generated using exploratory factor analysis, then tested with confirmatory factor analysis in the cross-diagnostic and SSD groups. The relationships between factor scores and computational speech features were examined for 202 of the participants. **Study Results:** We found a 3-factor model with a good fit in the cross-diagnostic group and an acceptable fit for the SSD subsample. The model identifies an *impaired expressivity* factor and 2 interrelated disorganized factors for *inefficient* and *incoherent* speech. Incoherent speech was specific to psychosis groups, while inefficient speech and impaired expressivity showed intermediate effects in people with nonpsychotic disorders. Each of the 3 factors had significant and distinct relationships with speech features, which differed for the cross-diagnostic v.s. SSD groups. **Conclusions:** We report a cross-diagnostic 3-factor model for speech disturbance which is supported by good statistical measures, intuitive, applicable to SSD, and relatable to linguistic theories. It provides a valuable framework for understanding speech disturbance and appropriate targets for modeling with quantitative speech features.

Key words: natural language processing/schizophrenia/psychosis/thought disorder/disorganization/alolia/graph analysis

Introduction

Quantitative features derived from the speech are increasingly recognized as valuable predictors and objective biomarkers for psychiatric disorders, notably including schizophrenia spectrum disorders (SSD).^{1–4} But, speech and language phenotypes in psychiatric disorders are heterogeneous. Which (and how many) clinical phenotypes should we target when modeling and deriving computational features? In this article, we identify cross-diagnostic latent factors for language disturbance based on clinical ratings, then demonstrate the relevance of these factors for computational linguistic modeling in relation to general psychopathology and in SSD. We regard speech as speech, without making inferences regarding “thought disorder,” a related construct that infers disruptions to thought based on observable changes in speech.

A range of speech features appears promising as predictors of psychiatric diagnoses and biomarkers of individual symptom dimensions. We use the term “speech features” broadly to indicate quantitative metrics derived from speech samples, including phonetic, acoustic, and textual measures. Significant advancements have been made in SSD to identify clinically relevant speech features. For example, SSD diagnosis can be classified with >80% accuracy relative to healthy volunteers (HV) using measurements of semantic distance, which quantify the “closeness” of the meaning in successive sentences or segments of words.^{5,6} This

strategy can be combined with other automated measurements of syntax/parts-of-speech,^{7–9} referential ambiguity,⁹ and metaphoricity.¹⁰ Acoustic features (pitch, voice quality, and pauses) also successfully predict SSD diagnosis.¹¹ Additionally, transition to psychosis among individuals at clinical high risk can be predicted with 83%–97% accuracy using automatic measurements of semantic density,¹² detections of metaphorical speech or nonstandard meanings,¹⁰ and a combination of semantic distance and parts-of-speech.^{13,14} Speech features are also promising biomarkers for other clinical contexts, including mania,¹⁵ depression,¹⁶ autism,¹⁷ and dementia.^{18,19}

Presumably, speech features predict clinical characteristics by reflecting speech-related symptoms such as decreased prosody, tangentiality, and changes in quantity. However, most studies have used speech features to directly predict diagnosis or outcomes, without relating features to observable speech disturbances.^{5–14} There are weaknesses to this approach. First, there is substantial heterogeneity in speech phenotypes within diagnoses, perhaps explaining the limited reproducibility of results across linguistic and experimental contexts.^{20,21} Cross-diagnostically, and within SSD, positive/disorganized vs negative/impoverished dimensions have been consistently reported^{2,22,23}; they are poorly and sometimes even negatively correlated with one another.²⁴ Therefore, greater precision may be achieved by modeling specific types of speech disturbance, rather than diagnostic phenotypes as a whole.^{24–26} Second, speech disturbances may be shared across disorders, with implications for underlying neurobiology. We need to define cross-diagnostically valid constructs in order to determine whether speech features can be used consistently across disorders as biomarkers for particular speech phenotypes. Notably, in a large sample ($N = 1071$), Stein et al. found 3 factors of speech disturbance (emptiness, disorganization, and incoherence) which were linked to changes in distinct brain regions and valid across diagnoses.²⁷ Speech graph features have been shown to be a good marker of thought disorder in mania and SSD,¹⁵ and to be related to cross-diagnostic psychopathological dimensions.²⁸ However, there are few studies examining whether quantitative speech features are consistently related to cross-diagnostic dimensions of speech disturbance.

Previous studies have reported factor analyses on clinical ratings for thought disorder, but results are varied, mostly limited to SSD cohorts, and relevance for computational modeling is unclear. The two-factor model distinguishing impoverished speech (eg, poverty of speech, latency, and concreteness) from disorganized speech (eg, tangentiality, derailment, incoherence, neologisms, and clanging) is the most consistent.^{2,22,23} However, “disorganized speech” is inconsistently defined across studies and encompasses a broad collection of individual symptoms which may require different computational strategies. A

6-factor model has been replicated in SSD, but several factors are not likely to be relevant cross-diagnostically (eg, idiosyncratic: Word approximations and stilted speech; referential: Echolalia and self-reference).^{23,29}

The objective of this study was to delineate dimensions of speech disturbance suitable as targets for computational modeling in SSD and across diagnoses. First, we used factor analyses to identify and test interpretable models based on clinical ratings for speech disturbances. Then, we evaluated the factors’ clinical relevance by comparing severity across diagnostic groups, with the expectation that clinically relevant factors should show different patterns in different groups. Finally, we related the factors to multimodal speech features cross-diagnostically and within an SSD subgroup, similarly to the expectation that for the factor solution to be computationally meaningful, the factors should be significantly and distinctly related to different speech features.

Methods

Participants

The cross-diagnostic sample included brief recordings of free speech (~1.5–3 min) from 343 individuals, including 90 definitively diagnosed with SSD (schizophrenia, schizoaffective disorder, schizophreniform disorder, delusional disorder, or unspecified psychotic disorder), 47 with affective psychosis (bipolar disorder or major depressive disorder with psychotic features) or probable but unconfirmed psychotic disorder (PSY), 130 with other nonpsychotic disorders (OD), and 76 HV (Table 1). Other psychiatric disorders included unipolar and bipolar mood disorders without psychotic features, anxiety disorders, obsessive-compulsive disorder, borderline personality disorder, attention-deficit hyperactivity disorder, and substance use disorders. Many participants were diagnosed with multiple comorbid conditions, so diagnoses were not mutually exclusive. Individuals with developmental, neurological, or medical conditions likely to impact speech were excluded, including intellectual disability, autism spectrum disorder, and dementia. Sampling approaches were heterogeneous, with 47 samples derived from publicly available videos of psychiatric interviews, 170 participants evaluated in person, and 126 participants evaluated virtually. All recruited participants signed informed consent, and human subjects research ethical approval was given by the IRB at the Feinstein Institutes for Medical Research. Additional details on ascertainment and sample collection are provided in the supplement.

Clinical Ratings of Speech Disturbance

To clinically characterize language disturbances, all samples were given ratings using the 18 items from the Scale for the Assessment of Thought Language and Communication (TLC)³⁰ and 2 additional speech-related

Table 1. Participant and Sample Characteristics

	Healthy Volunteers (HV)	Other Psychiatric Disorders (OD)	Other or Undetermined Psychosis (PSY)	Schizophrenia Spectrum Disorders (SSD)	Total Sample	<i>P</i> - Value
<i>N</i>	76	130	47	90	343	
Age - Mean (SD)	29.0 (6.7)	23.6 (5.4)	27.1 (8.5)	27.0 (6.7)	26.2 (6.7)	<.001
Gender - <i>n</i> (%)						.007
Man	35 (46%)	44 (34%)	25 (53%)	54 (60%)	155 (46%)	
Non-Binary	2 (3%)	11 (9%)	1 (2%)	4 (4%)	18 (5%)	
Woman	39 (51%)	74 (57%)	21 (45%)	30 (33%)	164 (48%)	
Race - <i>n</i> (%)						.003
Asian	11 (15%)	15 (14%)	4 (21%)	12 (14%)	42 (14%)	
Black	29 (38%)	19 (17%)	7 (37%)	37 (42%)	92 (31%)	
Multiple	5 (7%)	9 (8%)	0 (0%)	7 (8%)	21 (7%)	
Other	2 (2%)	6 (6%)	2 (11%)	11 (13%)	21 (7%)	
White	29 (38%)	60 (55%)	6 (32%)	20 (23%)	115 (40%)	
Unknown	0	21	28	3	52	
Hispanic - <i>n</i> (%)	10 (13%)	20 (18%)	2 (11%)	14 (16%)	46 (16%)	.72
Unknown	0	21	28	2	51	
Ethnicity						
TLC Global - mean (SD)	0.1 (0.4)	0.4 (0.6)	1.4 (1.2)	1.4 (1.0)	0.8 (1.0)	<0.001

Note: SD, standard deviation, TLC Global, Global score from the Scale for the Assessment of Thought Language and Communication.

items from the Scale for the Assessment of Negative Symptoms (SANS)³¹ which were not included in the TLC (SANS-06: Decreased Vocal Inflection; SANS-11: Increased Latency of Response). These scales were chosen because they are commonly used and represent a comprehensive range of psychiatrically relevant speech and language phenotypes. For prospectively evaluated participants, the ratings were based on the full assessment, including other language tasks and clinical interviews, and not specifically for the isolated sample that underwent computational analysis. Consistent with previous reports,^{23,30,32} we found low prevalence (absent in >90% of the sample), low interrater reliability, or low sampling adequacy for 6 TLC items: Echolalia, blocking, clanging, word approximation, self-reference, and stilted speech. These were not included in the analyses. Each of the remaining 14 items exhibited excellent interrater reliability (ICC≥0.9).

Speech Features

For a subsample of 202 participants (37 HV, 87 OD, 14 PSY, and 64 SSD), speech samples were transcribed verbatim and processed through an automated pipeline to extract acoustic (prosody and voice quality, speaking tempo, and pauses) and textual features (semantic distances, dysfluencies and speech errors, speech graph measures, lexical characteristics, sentiment, parts-of-speech, and speech quantity). This subsample includes only prospectively assessed participants responding to standardized open-ended prompts for whom data had

been fully processed at the time of the analysis. We initially selected 79 features for analysis that were previously found to be relevant for psychiatric disorders. To improve the interpretability of comparisons between computed speech features and clinical variables, we calculated variance inflation factors (VIF) for each category of features and omitted redundant features. Twenty-seven features were included in the final analysis. Additional details are provided in the supplement.

Factor Analyses

Exploratory factor analysis (eFA) was used to generate potential models for latent factors for clinical speech disturbance ratings in the cross-diagnostic sample, including all 343 participants. Usual assumptions were met: Bartlett Test of Sphericity: $P < .001$; Kaiser–Meyer–Olkin measure of sampling adequacy = 0.9; determinant = 0.0004. Visual inspection of the scree plot ([supplementary figure 1](#)) suggested 2–3 latent factors. We used the *psych* package³³ in R³⁴ to generate 2- and 3-factor principal axis solutions with Promax rotations. We chose an oblique rotation because we hypothesized that latent factors may be correlated with one another. Confirmatory factor analysis (cFA) was used to examine fit statistics for the full cross-diagnostic sample and the SSD subgroup. Maximum likelihood estimation from *lavaan* package³⁵ was used. Speech factor scores were computed based on the final 3-factor model and the main group effect on factor scores was evaluated using ANOVA and pair-wise comparisons were made using t-tests.

Correlations and Network Generation

Spearman correlations were calculated between factor scores and computed speech features. A network was constructed to further illustrate correlational relationships. Nodes represent factor scores, speech features, and clinical characteristics. Edges represent Spearman correlation coefficients with $\rho > 0.2$ and $P < .05$. The graph was plotted in R using the *igraph* package³⁶ and network descriptors were calculated for the degree (number of connections for each node), betweenness centrality (how much the node serves as a conduit for the shortest paths connecting all node-pairs), and the overall density of the graph (interconnectedness of all nodes). All metadata, code, and resources for replicating our factor score calculations are available at: <https://github.com/STANG-lab/Analysis/tree/main/Factor-network>.

Accounting for Demographic and Sampling Differences

There were significant group differences in gender, age, and race as well as the sampling method. To account for the demographic group differences, we repeated the cFA in the following subsamples: matching for gender ($N = 188$), including all 4 groups; matching for age and race ($N = 228$), excluding the PSY group due to insufficient samples. Of note, age remained statistically significantly different but averaged in the mid-late twenties for all groups. We also reexamined correlations between factor scores and speech features in another subsample matched for age, gender, and race ($N = 111$). Optimal matching with *probit* propensity scores was completed using the *MatchIt* package in R.³⁷ Because different sampling strategies were used for different participant populations (eg, inpatients participated in person, outpatients, and HV generally participated virtually), it was not possible to take a matched-sample approach. For the factor analyses, the purpose was to identify latent factors which explain how speech and language disturbances manifest, as reflected by subjective clinical ratings—ie, which types of symptoms tend to occur together. Therefore, we optimize the accuracy and generalizability of the results by including the largest and most diverse sample possible, and sampling variability is not a major limitation as long as the ratings are done consistently and reliably. Our purpose in looking at speech features was to validate the factor scores as making meaningful distinctions in observable speech and language disturbance which may generate better, targeted analytical methods and more accurate computational modeling of speech and language disturbance down the road. Modeling these factors with specific features was not within the scope of this article and may require a series of iterative future efforts. However, even in this limited scope, sampling method may significantly affect computed speech features. We limit this by including only samples from prospectively collected open-ended prompts (albeit some in person

and some virtually, with different recording devices), and choosing normalized acoustic features which should not be affected by recording method. We also include an additional correlational analysis which teases apart assessment methods.

Results

Latent Factors of Language Disturbance

Results of the eFA (table 2, supplementary figures 2 and 3) suggested 2- and 3-factor models based on 14 clinical ratings for speech and language disturbance symptoms in the cross-diagnostic sample. The 2-factor model identified factors related to *disorganized speech* and *impaired expressivity* (decreased speech content and expressiveness), explaining 40% and 12% of the variance, respectively. The 3-factor model also produced the *impaired expressivity* factor (5% variance) and further divided disorganized speech into items consistent with *inefficient speech* (poor organization across ideas; 41% of variance) and *incoherent speech* (nonsensical or unintelligible utterances; 12% of variance). Each model was tested using cFA in both the overall sample and the SSD subgroup. The 2-factor model was a poor fit for both the cross-diagnostic and SSD samples (Cross-diagnostic: Comparative Fit Index (CFI) = 0.850, Tucker-Lewis Index (TLI) = 0.821, Root Mean Square Error of Approximation (RMSEA) = 0.095; SSD: CFI = 0.832, TLI = 0.799, RMSEA = 0.117). Because the 3-factor model had multiple cross-loadings which are not suitable for cFA, we tested this model in 2 ways: First, by omitting the cross-loaded items (Poverty of Content of Speech, Derailment, and Loss of Goal) and then by including a separate fourth factor with the cross-loaded items. Both approaches demonstrated good fit for the final model in the cross-diagnostic sample: Without cross-loaded items (3-factor): CFI = 0.965, TLI = 0.953, RMSEA = 0.047; With separate factor (4-factor): CFI = 0.954, TLI = 0.941, RMSEA = 0.054. In the SSD subsample, the fit was acceptable without cross-loaded items (3-factor): CFI = 0.911, TLI = 0.880, RMSEA = 0.085; and good with the separate factor (4-factor): CFI = 0.918, TLI = 0.923, RMSEA = 0.074.

Group Differences in Factor Scores

Each of the 3 factors exhibited significant group effects ($P < .001$) but with different patterns (figure 1). Inefficient speech showed a graded effect with significant differences between each pair of groups; inefficient speech was highest in the SSD group, followed by PSY, then OD, and HV. Incoherent speech was specific to psychosis and was elevated in both psychotic groups (SSD and PSY) compared to both nonpsychotic groups (OD and HV); with no significant difference between SSD and PSY or between OD and HV. Clinically significant impaired expressivity symptoms were present across all groups, but

Table 2. Factor Loadings

2-Factor Model		Speech and Language Items	3-Factor Model		
Disorganized Speech	Impaired Expressivity		Inefficient Speech	Incoherent Speech	Impaired Expressivity
	0.77	Poverty of Speech (TLC-01)			0.81
0.87		Poverty of Content of Speech (TLC-02)	0.46	0.47	
0.65	-0.34	Pressured Speech (TLC-03)	0.79		
0.49		Distractible Speech (TLC-04)	0.58		
0.85		Tangentiality (TLC-05)	0.92		
0.89		Derailment (TLC-06)	0.58	0.39	
0.65		Incoherence (TLC-07)		0.89	
0.73		Illogicality (TLC-08)		0.67	
0.38		Neologism (TLC-10)		0.58	
0.72		Circumstantiality (TLC-12)	0.84		
0.82		Loss of Goal (TLC-13)	0.45	0.43	
0.66		Perseverations (TLC-14)	0.58		
	0.62	Decreased Vocal Inflections (SANS-06)			0.69
	0.57	Increased Latency of Response (SANS-11)			0.56

Note: Loadings <0.3 are masked. TLC, Scale for the Assessment of Thought Language and Communication; SANS, Scale for the Assessment of Negative Symptoms.

highest in the psychosis groups (SSD and PSY), intermediate for participants with other psychiatric disorders (OD), and lowest for healthy volunteers (HV). The group effects were consistent when examining subsamples where we attempted to match for gender, age, and race (supplementary figures 4 and 5).

Relationships to Computational Features

Figure 2 illustrates the distinct relationships between factor scores and computed speech features. For example, in the SSD group, inefficient speech was most closely related to speaking rate ($\rho = 0.63$) and mean age of acquisition ($\rho = 0.54$); incoherent speech was most closely related to decreased positive valence ($\rho = -0.50$) and use of particles ($\rho = 0.41$), along with mean speaking rate ($\rho = 0.41$); and impaired expressivity was most closely related to decreased mean pitch ($\rho = -0.41$) and higher mean turn latency ($\rho = 0.41$). The strength of the correlations, and occasionally, the direction, also differed across the 4 groups. However, in all cases, none of the 3 factors showed redundant relationships to computed speech features. This was also true in the demographically matched subsample (supplementary table 7 and figure 6). Examining assessment types separately, there is also variability between SSD inpatients assessed in person and SSD outpatients assessed virtually (supplementary figure 7).

Networks generated for the factor scores further demonstrate these distinct relationships with speech features, as well as interrelationships among all measures for the cross-diagnostic and SSD samples (figure 3). The overall densities of the networks were similar (cross-diagnostic: 0.21, SSD: 0.19), suggesting that there was a similar amount of interconnectedness among all features. In both cases, among the factor scores, inefficient speech showed the greatest amount of connectedness to speech features as reflected by degree (cross-diagnostic: 8; SSD: 9) and betweenness centrality (cross-diagnostic: 25.3; SSD: 47.9). In the cross-diagnostic sample, incoherence was the least connected factor, while in the SSD sample, impaired expressivity was least connected, perhaps reflecting the extent to which variability in speech and language phenotype is related to these factors. The relationships to computed speech features differed for each of the 3-factor scores, and also between the cross-diagnostic and SSD samples, as further illustrated through dendrograms in supplementary figure 8.

Discussion

In this study, we identified a 3-factor model that describes speech and language disturbance with good fit in cross-diagnostic and SSD samples. An *impaired expressivity* factor included items related to decreased quantity and

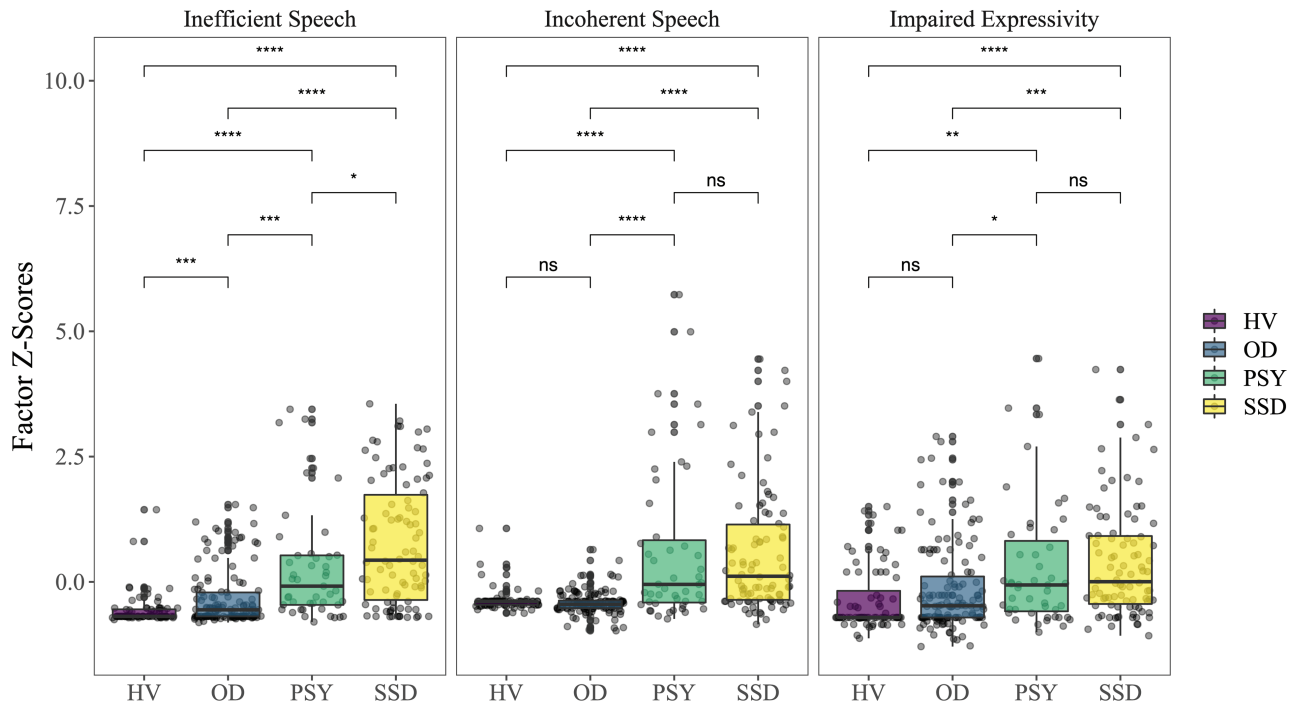


Fig. 1. Group Differences In Factor Scores: Pair-wise comparisons made using t-tests. Significance levels shown on the graph are uncorrected values. Results are largely consistent after correcting for multiple comparisons with the FDR method, the pair-wise P values are as follows. Inefficient speech: HV*OD = 0.05; PSY*SSD = 0.003; all others $P < .001$. Incoherent speech: HV*OD & PSY*SSD NS; all others $P < .001$. Impaired Expressivity: HV*OD & PSY*SSD NS; HV*OD, HV*SSD, and OD*SSD $P < .001$; OD*PSY = 0.01. NS, Not significant; * $P < .05$; ** $P < .01$; *** $P < .001$. HV, healthy volunteers; OD, other psychiatric disorders; PSY, other or undetermined psychotic disorders; SSD, schizophrenia spectrum disorders.

expressiveness. Two interrelated disorganized factors emerged: an *inefficiency* factor included items describing poorly related, redundant, or excessive speech, and an *incoherent* factor included items relating to nonsensical or unintelligible speech. Fittingly, derailment, loss of goal, and poverty of content of speech were cross-loaded on both the inefficient and the incoherent factors. Our confirmatory models suggest a good fit for the 3-factor model in the cross-diagnostic sample, and an adequate fit for the SSD subgroup which is comparable to previously reported models.²³ The distinction between impaired expressivity and disorganization-type symptoms is well-replicated in SSD²³ and cross-diagnostically.² A latent factor for the poverty of thought and decreased expressiveness has been found by independent groups and using other rating scales.^{38,39} Factor analysis of the Thought and Language Index also suggests that unusual word usage, sentence structure, and logic (consistent with our *incoherence* factor) may be differentiated from distractibility and perseveration (included in our *inefficiency* factor).⁴⁰ Overall, our model is very similar to the 3-factor model for formal thought disorder reported by Stein et al. in their cross-diagnostic sample of $N = 1071$.²⁷ They describe an emptiness factor (poverty of speech and content, increased latency, and blocking), a disorganization factor (tangentiality, circumstantiality, derailment, and pressure of speech), and an incoherence

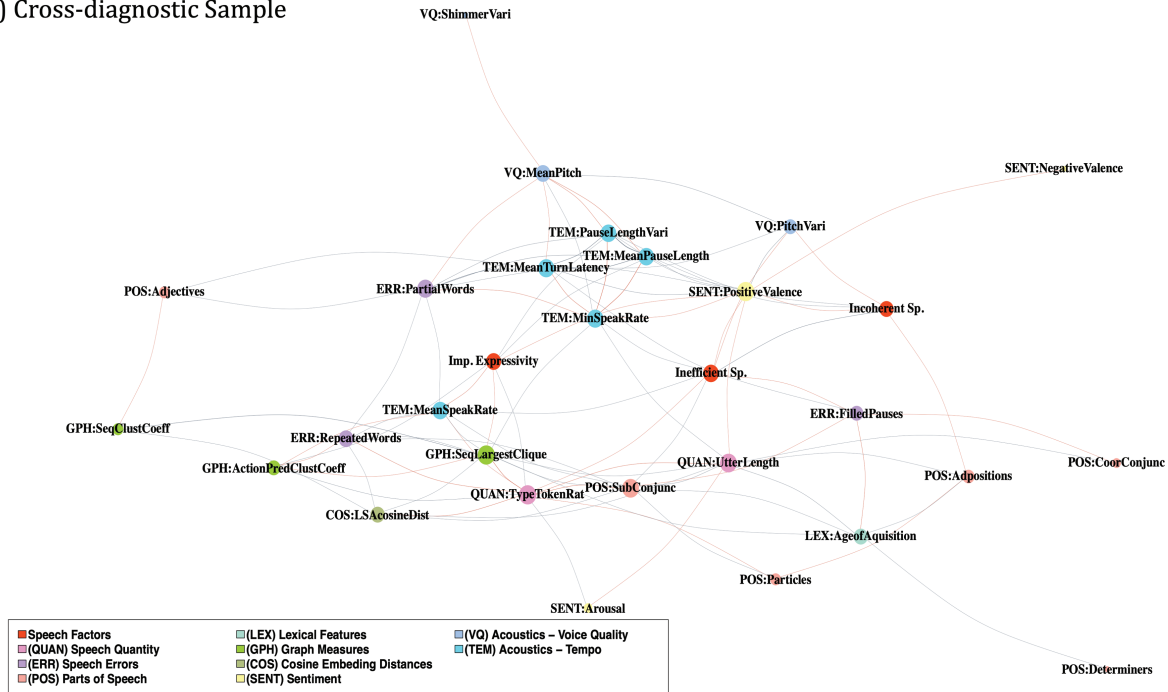
factor (incoherence, illogicality, and distractibility). The principal differences are that we include decreased vocal inflection in our impaired expressivity factor (since we are looking at speech as a whole, and not specifically thought disorder) and that we prefer the term “inefficiency” over “disorganization” because “disorganization” is an overly broad term that can include incoherence and other items. Notably, Stein et al. report distinct correlations to brain structure for their 3 factors. The similarity between our findings, and the fact that the samples were collected in different languages and rated by different teams, provides added confidence for the 3-factor model we propose.

The 3-factor model proposed here can be understood in the context of linguistic theories on pragmatics. In classical Gricean pragmatics, a speech act is carried out successfully when the addressee is made to recognize the speaker’s communicative intent (“Meaning-intention” or “M-intention”).^{41,42} Understanding the M-intention necessitates not only deciphering the semantic content (literal meaning) of what is spoken, but also the conversational implicatures which are grounded in the context of the discourse and in the cooperative principle. Per Grice, the cooperative principle assumes that a cooperative and rational speaker is always following 4 maxims: (1) Quantity: Be informative, but not overly informative, (2) Quality: Be truthful, to the best of one’s knowledge, (3) Relation: Be relevant, and (4) Manner: Be perspicuous

	All N=202			HV N=37			OD N=87			PSY N=14			SSD N=64		
	Inefficient Sp.	Incoherent Sp.	Imp. Expressivity	Inefficient Sp.	Incoherent Sp.	Imp. Expressivity	Inefficient Sp.	Incoherent Sp.	Imp. Expressivity	Inefficient Sp.	Incoherent Sp.	Imp. Expressivity	Inefficient Sp.	Incoherent Sp.	Imp. Expressivity
QUAN:TypeTokenRat	-0.23	-0.04	0.23	-0.3	-0.23	0.56	-0.29	0.06	-0.07	-0.21	0.18	0.67	-0.29	-0.25	0.03
QUAN:UtterLength	0.15	-0.05	-0.13	0.03	-0.12	-0.19	0.42	-0.08	0.19	0.3	-0.06	-0.65	0.19	0.18	-0.06
GPH:SeqClustCoeff	0.09	0.03	0.1	0.35	0.09	-0.16	-0.12	0.2	0.22	-0.43	0.45	-0.03	0.24	-0.09	0.07
GPH:SeqLargestClique	0.28	0.06	-0.2	0.3	0.25	-0.48	0.25	-0.06	-0.03	0.24	0.18	-0.61	0.51	0.29	-0.15
GPH>ActionPredClustCoeff	-0.14	0	0.29	-0.24	-0.04	0.23	-0.29	-0.07	0.19	-0.03	0.23	0.45	-0.12	-0.06	-0.01
POS:Adjectives	-0.07	0.1	-0.06	-0.38	-0.22	0.21	0.11	0.05	-0.02	-0.1	0.23	-0.35	-0.08	0.07	-0.06
POS:Adpositions	-0.09	-0.08	-0.14	-0.01	0.04	-0.03	-0.03	-0.06	-0.01	-0.22	0.19	0.02	0	-0.01	0
POS:CoorConjunc	0.04	0.01	0.05	0.22	-0.2	-0.09	-0.08	-0.18	0.18	-0.04	0.52	0.17	0.15	0.17	0.1
POS:SubConjunc	-0.03	0.07	0.09	-0.22	-0.03	0.17	0.11	0.13	0.23	0.34	0.1	-0.03	-0.03	0.05	0.1
POS:Determiners	0.07	0.04	0.06	-0.08	-0.06	-0.2	0.26	0.04	0.09	-0.62	-0.08	0.17	-0.04	-0.06	-0.05
POS:Particles	0.19	0.16	-0.09	-0.02	0.14	-0.2	0.27	0.16	-0.02	0.53	-0.17	-0.47	0.22	0.41	0.09
ERR:PartialWords	-0.07	0.2	-0.07	0	0.33	0.08	0.05	0.25	-0.05	-0.09	0.16	-0.51	-0.1	0	0.19
ERR:RepeatedWords	0.08	0	-0.04	0.21	0.04	0.06	0.11	0.03	-0.07	-0.31	-0.06	-0.25	0.22	0.02	0.11
ERR:FilledPauses	-0.26	-0.03	0.1	-0.1	0.19	0.26	-0.28	0.04	0.14	-0.34	-0.36	-0.06	-0.51	-0.26	0.15
COS:LSAcosineDist	0.1	-0.11	-0.09	0.16	-0.08	-0.26	0.07	-0.11	-0.03	-0.08	0.1	-0.64	0.22	-0.14	-0.05
LEX:AgeofAquisition	0.21	0.02	-0.12	0.08	-0.22	-0.27	0.08	-0.04	0.09	0.41	0.24	-0.48	0.54	0.27	-0.07
SENT:PositiveValence	-0.18	-0.1	-0.09	0.03	-0.09	-0.02	-0.2	0.09	-0.05	-0.04	-0.23	0.22	-0.33	-0.5	-0.01
SENT:NegativeValence	-0.09	-0.03	-0.1	-0.14	0.07	0.23	0.07	-0.15	-0.1	-0.01	0.15	-0.17	-0.08	0.18	0.01
SENT:Arousal	-0.23	-0.03	0.08	-0.29	-0.02	0.48	-0.26	0.19	-0.03	0.31	0.31	0.36	-0.16	-0.24	0.03
VQ:MeanPitch	0.07	-0.14	-0.1	0.32	0.15	-0.39	0.01	-0.33	0.02	-0.12	0.15	0.63	-0.05	-0.15	-0.41
VQ:PitchVari	-0.16	-0.14	-0.05	0.16	0.13	-0.18	-0.11	-0.04	0.19	-0.11	0.37	0.73	-0.22	-0.39	0.04
VQ:ShimmerVari	-0.14	-0.06	-0.24	-0.02	-0.05	-0.24	-0.18	0.04	-0.04	0.38	0.22	-0.27	-0.29	-0.28	-0.29
TEM:MeanPauseLength	-0.15	0.22	0.18	-0.25	0.07	0.16	-0.25	0.3	0.4	-0.27	0.34	-0.06	-0.15	-0.05	0.08
TEM:PauseLengthVari	-0.12	0.19	0.19	-0.2	0.03	0.23	-0.19	0.24	0.37	-0.27	0.3	-0.08	-0.13	-0.13	0.17
TEM:MeanSpeakRate	0.33	0.09	-0.42	0.56	0.04	-0.44	0.21	-0.06	-0.49	0.49	0.04	-0.75	0.63	0.41	-0.15
TEM:MinSpeakRate	0.18	-0.11	-0.12	0.08	0.21	-0.37	0.28	-0.27	-0.11	0.36	-0.41	-0.1	0.15	0.11	-0.36
TEM:MeanTurnLatency	-0.07	0.19	0.02	-0.23	0.13	0.23	-0.21	0.01	-0.04	0.36	0.08	-0.23	-0.15	0.07	0.41

Fig. 2. Spearman Correlations between Factor Scores and Computed Speech Features. HV, healthy volunteers; OD, other psychiatric disorders; PSY, other or undetermined psychotic disorders; SSD, schizophrenia spectrum disorders. See [supplementary table 3](#) for descriptions of the speech features.

A) Cross-diagnostic Sample



B) Schizophrenia Spectrum Disorders

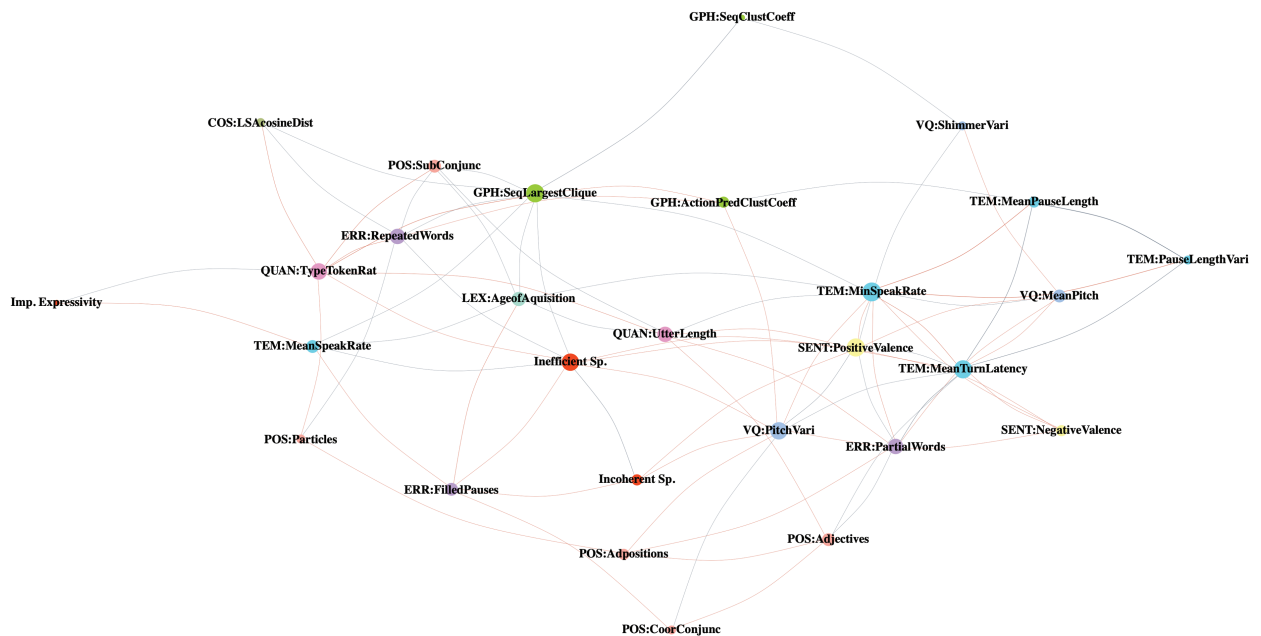


Fig. 3. Network of Factor Scores and Computed Speech Features: Nodes represent factor scores, acoustics, and lexical features; size is proportional to the degree of the node. Edges represent Spearman correlation coefficients with cutoff of $\rho = 0.2$ and $P = .05$; weight is proportional to absolute value. A) Cross-diagnostic sample ($n = 202$). Density = 0.21. Degree (D) of each speech factor: Inefficient Speech = 8, Incoherent Speech = 6, Impaired Expressivity = 7. Betweenness centrality (BC) of each speech factor: Inefficient Speech = 25.3, Incoherent Speech = 13.0, Impaired Expressivity = 18.3. Speech features with the highest connectedness: type-token ratio (D = 11, BC = 31.8), largest clique in sequential graph (D = 11, BC = 42.0), positive sentiment (D = 11, BC = 49.6), mean sentence length (D = 10, BC = 63.8). B) Schizophrenia Spectrum Disorders ($n=64$). Density = 0.18. Degree of each speech factor: Inefficient Speech = 9, Incoherent Speech = 4, Impaired Expressivity = 2. Betweenness centrality of each speech factor: Inefficient Speech = 47.9, Incoherent Speech = 2.4, Impaired Expressivity = 0.37. Speech features with the highest connectedness: minimum speaking rate (D=11, BC=50.1), largest clique in sequential graph (D=10, BC=54.4), positive sentiment (D = 10, BC = 33.4), mean turn latency (D = 10, BC = 29.1), pitch range (D = 9, BC = 35.5).

and avoid ambiguity (clear). Along these lines, we can characterize the impaired expressivity factor as violation of the maxim on quantity—specifically, delivery of insufficient information. Inefficiency and incoherence both involve violations of relation, manner, and (excessive) quantity, but differ in degree. With incoherent speech, the violations are severe to the point of preventing the M-intention from being inferred. With inefficient speech, the violations of the cooperative principle are perceived, but the M-intention can still be ascertained, although costly pragmatic reasoning might be needed to make sense of the utterance. We can also differentiate between incoherent and inefficient disorganized speech by referring to Grosz's theories on centering and discourse structure.^{43,44} She suggests that, just as individual sentences can be broken down into a hierarchical structure of phrases that give meaning to one another, so can discourse be broken into a hierarchy of discourse segments (each containing one or more utterances). In coherent discourse, individual utterances within a discourse segment collaborate to convey the discourse segment purpose (local coherence), and multiple discourse segments relate to one another to satisfy the overall high-level discourse purpose (global coherence). With respect to our work, we would characterize inefficiency as disruptions in global coherence, where there are disruptions or inefficiencies in satisfying the overall discourse purpose, while incoherence arises from disruptions to local coherence that cause the discourse segment purpose to be obscured.

We also found distinct and significant relationships to different diagnostic groups for each of the 3 factors, suggesting that the model is not just statistically and theoretically sound, but also clinically meaningful. There were intermediate effects for nonpsychotic disorders in inefficiency and impaired expressivity, but incoherence was specific to people with psychosis and rarely elevated in either HV or people with nonpsychotic disorders. Intuitively, this pattern may be explained by the sensitivity of impaired expressivity and inefficiency to nonpsychiatric (e.g. personality, culture, and social context)^{45,46} and nonspecific variables (eg, impaired attention, psychomotor retardation, and ruminations and repetitive thinking).^{27,47} In contrast, incoherence may be more specific to psychosis-related brain changes.

The objective of examining the relationships between factor scores and computed speech features was to examine the relevance of the factors for computational speech and language analysis. The expectation is that meaningful distinctions in clinical speech dimensions should correspond to nonredundant relationships to different speech features; i.e. that different computational strategies should be pursued to objectively characterize each of these dimensions of speech disturbance. Indeed, across all diagnostic groups and sampling strategies, the 3 factors showed distinct relationships to computed speech features. Some of the relationships recapitulate findings in

the literature: eg, “negative symptoms” are related to decreased speaking rate.^{48,49} Other findings need to be more closely examined to tease apart diagnostic, demographic, and sampling effects. We believe that the distinction between inefficient and incoherent speech is particularly important to note with respect to computational analyses because departing from previous works which usually conflate the 2, we can imagine that some strategies are better suited to target one vs the other—e.g. perhaps calculating distances better measures inefficiency and is applicable cross-diagnostically, while perplexity paradigms may better measure incoherence and be more specific to psychosis-related language disturbance. It is possible that focusing on cross-diagnostically relevant symptom dimensions as represented by the factors may improve the reproducibility and generalizability of findings relating computational speech features to psychopathology.^{20,21}

There were several limitations in our work. There were heterogeneities in demographic characteristics and data collection methods, which require further validation in independent samples to fully resolve. We attempted to limit these concerns by evaluating demographically matched subsamples, looking only at open-ended autobiographical narrative speech, and choosing speech features that should be minimally affected by different recording situations. We were not fully able to match for age in the sample examining group effects on the factor scores, but the absolute differences were small. Because different sampling strategies were used for different participant groups, not all of the potential moderating effects can be teased apart—for example, the effects of in-person vs virtual assessment. In addition, while we attempted to standardize recording conditions by conducting our sessions in a private space with minimal background noise, we did not systematically evaluate or correct for background noise. All speech samples used for the computational analyses were in response to 2 autobiographical narrative questions, and participants were encouraged to speak for at least 2 minutes, but there was still significant variability in the amount of speech produced, possibly influenced by the assessment environment (inpatient vs outpatient and in-person vs virtual). We attempted to normalize for this variability based on the total number of words or utterances, where appropriate. While brief assessments may be more efficient, they may also be less sensitive or accurate than longer language samples, which allow for symptoms to emerge and evolve. Additionally, all samples were collected in North American English and rated with the TLC scale (with 2 SANS items) at a single site. A larger multilingual study using multiple rating scales would further support the generalizability of our findings. Here, we were interested in the real-world phenomenology of speech and language symptoms, so we did not account for medication effects which are likely to have a significant effect on both clinical ratings and computational features. Future studies should investigate

the relative impact of medications vs underlying psychopathology on speech and language symptoms.

In this study, we report a cross-diagnostic 3-factor model for speech and language disturbance which is supported by good statistical measures, intuitive, applicable to SSD, and relatable to linguistic theories. *Impaired expressivity, inefficiency, and incoherence* show meaningfully distinct patterns in different diagnostic groups, with incoherent speech being most specific to psychosis. Each factor was closely but distinctly related to other clinical characteristics and computational speech features. The factors also inspire different computational strategies, perhaps allowing for improved accuracy, specificity, and reproducibility relative to overly heterogeneous diagnostic groups or overly specific individual symptoms. In conclusion, the 3-factor model reported here is a valuable framework for understanding speech and language disturbance cross-diagnostically and in SSD particularly, and the factor scores are appropriate targets for modeling with quantitative speech features.

Supplementary Material

Supplementary material is available at <https://academic.oup.com/schizophreniabulletin/>.

Acknowledgments

This project was supported by the Brain and Behavior Research Foundation Young Investigator Award (SXT) and the American Society of Clinical Psychopharmacology Early Career Research Award (SXT). Data for a portion of the participants ($n = 210$) was collected in partnership with, and with financial support from, Winterlight Labs, but the conceptualization for this project, computation of speech features, and analyses were completed independently. We thank the participants for their contributions. We also thank Danielle DeSouza, Bill Simpson, Jessica Robin, and Liam Kaufman of Winterlight Labs for their ongoing collaboration. We are grateful to the clinicians and leaders at Zucker Hillside Hospital for their support of our work, including Drs. Michael Birnbaum, Anna Costakis, Ema Saito, Anil Malhotra, and John Kane. We thank James Fiumara and Jonathan Wright from the Linguistic Data Consortium, and Aamina Dhar, Grace Serpe, Jessica Guo and Styliana Maimos for their help in the transcription process.

Conflicts of Interest

SXT is a consultant for Neurocrine Biosciences and North Shore Therapeutics, received funding from Winterlight Labs, and holds equity in North Shore Therapeutics. The other authors have no conflicts of interest.

References

1. Corcoran CM, Mittal VA, Bearden CE, *et al.* Language as a biomarker for psychosis: a natural language processing approach. *Schizophr Res.* 2020;226:158–166. doi: [10.1016/j.schres.2020.04.032](https://doi.org/10.1016/j.schres.2020.04.032).
2. Kircher T, Bröhl H, Meier F, Engelen J. Formal thought disorders: from phenomenology to neurobiology. *Lancet Psychiatry.* 2018;5(6):515–526. doi: [10.1016/S2215-0366\(18\)30059-2](https://doi.org/10.1016/S2215-0366(18)30059-2).
3. de Boer JN, Brederoo SG, Voppel AE, Sommer IEC. Anomalies in language as a biomarker for schizophrenia. *Curr Opin Psychiatry.* 2020;33(3):212–218. doi: [10.1097/YCO.0000000000000595](https://doi.org/10.1097/YCO.0000000000000595).
4. Ratana R, Sharifzadeh H, Krishnan J, Pang S. A comprehensive review of computational methods for automatic prediction of schizophrenia with insight into indigenous populations. *Front Psychiatry.* 2019;10(September):1–15. doi: [10.3389/fpsy.2019.00659](https://doi.org/10.3389/fpsy.2019.00659).
5. Voppel A, de Boer J, Brederoo S, Schnack H, Sommer I. Quantified language connectedness in schizophrenia-spectrum disorders. *Psychiatry Res.* 2021;304:114130. doi: [10.1016/j.psychres.2021.114130](https://doi.org/10.1016/j.psychres.2021.114130).
6. Sarzynska-Wawer J, Wawer A, Pawlak A, *et al.* Detecting formal thought disorder by deep contextualized word representations. *Psychiatry Res.* 2021;304:114135. doi: [10.1016/j.psychres.2021.114135](https://doi.org/10.1016/j.psychres.2021.114135).
7. Tang SX, Kriz R, Cho S, *et al.* Natural language processing methods are sensitive to sub-clinical linguistic differences in schizophrenia spectrum disorders. *NPJ Schizophr.* 2021;7:25.
8. 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(3):270–284. doi: [10.1016/j.jneuroling.2009.05.002](https://doi.org/10.1016/j.jneuroling.2009.05.002).
9. Iter D, Yoon J, Jurafsky D. Automatic detection of incoherent speech for diagnosing schizophrenia. In: *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic.* Association for Computational Linguistics; 2018:136–146. doi: [10.18653/v1/W18-0615](https://doi.org/10.18653/v1/W18-0615).
10. Gutiérrez ED, Corlett PR, Corcoran CM, Cecchi GA. Using automated metaphor identification to aid in detection and prediction of first-episode schizophrenia. Association for Computational Linguistics. 2017 Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (January 2018):2923–2930. doi: [10.18653/v1/d17-1316](https://doi.org/10.18653/v1/d17-1316).
11. de Boer JN, Voppel AE, Brederoo SG, *et al.* Acoustic speech markers for schizophrenia-spectrum disorders: a diagnostic and symptom-recognition tool. *Psychol Med.* 2021:1–11. doi: [10.1017/S0033291721002804](https://doi.org/10.1017/S0033291721002804).
12. Rezaii N, Walker E, Wolff P. A machine learning approach to predicting psychosis using semantic density and latent content analysis. *NPJ Schizophr.* 2019;5(1):9. doi: [10.1038/s41537-019-0077-9](https://doi.org/10.1038/s41537-019-0077-9).
13. Bedi G, Carrillo F, Cecchi GA, *et al.* Automated analysis of free speech predicts psychosis onset in high-risk youths. *NPJ Schizophr.* 2015;1(1):15030. doi: [10.1038/npjjschz.2015.30](https://doi.org/10.1038/npjjschz.2015.30).
14. Corcoran CM, Carrillo F, Fernández-Slezak D, *et al.* Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry.* 2018;17(1):67–75. doi: [10.1002/wps.20491](https://doi.org/10.1002/wps.20491).
15. Mota NB, Furtado R, Maia PPC, Copelli M, Ribeiro S. Graph analysis of dream reports is especially informative

- about psychosis. *Sci Rep*. 2015;4(1):3691. doi: [10.1038/srep03691](https://doi.org/10.1038/srep03691).
16. Jiang H, Hu B, Liu Z, et al. Detecting depression using an ensemble logistic regression model based on multiple speech features. *Comput Math Methods Med*. 2018;2018:1–9. doi: [10.1155/2018/6508319](https://doi.org/10.1155/2018/6508319).
 17. Parish-Morris J, Liberman M, Ryant N, et al. Exploring autism spectrum disorders using HLT. *Proceedings of the 3rd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. 2016:74–84. doi: [10.18653/v1/w16-0308](https://doi.org/10.18653/v1/w16-0308).
 18. Yeung A, Iaboni A, Rochon E, et al. Correlating natural language processing and automated speech analysis with clinician assessment to quantify speech-language changes in mild cognitive impairment and Alzheimer's dementia. *Alzheimers Res Ther*. 2021;13(1):109. doi: [10.1186/s13195-021-00848-x](https://doi.org/10.1186/s13195-021-00848-x).
 19. Luzzi S, Baldinelli S, Ranaldi V, et al. The neural bases of discourse semantic and pragmatic deficits in patients with frontotemporal dementia and Alzheimer's disease. *Cortex*. 2020;128:174–191. doi: [10.1016/j.cortex.2020.03.012](https://doi.org/10.1016/j.cortex.2020.03.012).
 20. Parola A, Lin JM, Simonsen A, et al. Speech disturbances in schizophrenia: assessing cross-linguistic generalizability of NLP automated measures of coherence. *Schizophr Res*. 2022;S0920996422002742. doi: [10.1016/j.schres.2022.07.002](https://doi.org/10.1016/j.schres.2022.07.002).
 21. Parola A, Lin JM, Simonsen A, et al. Speech disturbances in schizophrenia: assessing cross-linguistic generalizability of NLP automated measures of coherence. *Schizophr Res*. 2022;S0920996422002742. doi: [10.1016/j.schres.2022.07.002](https://doi.org/10.1016/j.schres.2022.07.002).
 22. Kircher T, Krug A, Stratmann M, et al. A rating scale for the assessment of objective and subjective formal Thought and Language Disorder (TALD). *Schizophr Res*. 2014;160(1-3):216–221. doi: [10.1016/j.schres.2014.10.024](https://doi.org/10.1016/j.schres.2014.10.024).
 23. Cuesta MJ, Peralta V. Thought disorder in schizophrenia. Testing models through confirmatory factor analysis. *Eur Arch Psychiatry Clin Neurosci*. 1999;249(2):55–61. doi: [10.1007/s004060050066](https://doi.org/10.1007/s004060050066).
 24. Hitzzenko K, Mittal VA, Goldrick M. Understanding language abnormalities and associated clinical markers in psychosis: the promise of computational methods. *Schizophr Bull*. 2020;1:19. doi: [10.1093/schbul/sbaa141](https://doi.org/10.1093/schbul/sbaa141).
 25. Krell R, Tang W, Hänsel K, Sobolev M, Cho S, Tang SX. Lexical and Acoustic correlates of clinical speech disturbance in schizophrenia. *W3PHAI 2021 Studies in Computational Intelligence*. 2021;1013:9. doi: [10.1007/978-3-030-93080-6_3](https://doi.org/10.1007/978-3-030-93080-6_3).
 26. Haas SS, Doucet GE, Garg S, et al. Linking language features to clinical symptoms and multimodal imaging in individuals at clinical high risk for psychosis. *Eur Psychiatry*. 2020;63(1):e72. doi: [10.1192/j.eurpsy.2020.73](https://doi.org/10.1192/j.eurpsy.2020.73).
 27. Stein F, Buckenmayer E, Brosch K, et al. Dimensions of formal thought disorder and their relation to gray- and white matter brain structure in affective and psychotic disorders. *Schizophr Bull*. 2022;48:sbac002. doi: [10.1093/schbul/sbac002](https://doi.org/10.1093/schbul/sbac002).
 28. Nikzad AH, Cong Y, Berretta S, et al. Who does what to whom? graph representations of action-predication in speech relate to psychopathological dimensions of psychosis. *Nature Schizophrenia*. 2022;8:58. doi: [10.1101/2022.03.18.22272636](https://doi.org/10.1101/2022.03.18.22272636).
 29. Peralta V, Cuesta MJ, de Leon J. Formal thought disorder in schizophrenia: a factor analytic study. *Compr Psychiatry*. 1992;33(2):105–110. doi: [10.1016/0010-440X\(92\)90005-B](https://doi.org/10.1016/0010-440X(92)90005-B).
 30. Andreasen NC. Scale for the assessment of thought, language, and communication (TLC). *Schizophr Bull*. 1986;12(3):473–482. doi: [10.1093/schbul/12.3.473](https://doi.org/10.1093/schbul/12.3.473).
 31. Andreasen NC. *Scale for the Assessment of Negative Symptoms (SANS)*. Iowa City, Iowa: University of Iowa; 1984.
 32. Andreasen NC, Grove WM. Thought, language, and communication in schizophrenia: diagnosis and prognosis. *Schizophr Bull*. 1986;12(3):348–359. doi: [10.1093/schbul/12.3.348](https://doi.org/10.1093/schbul/12.3.348).
 33. Revelle W. *psych: Procedures for Psychological, Psychometric, and Personality Research*. Illinois: Northwestern University in Evanston; 2021. <https://CRAN.R-project.org/package=psych>.
 34. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2020. <https://www.R-project.org/>.
 35. Rosseel Y. lavaan: AN R package for structural equation modeling. *J Stat Soft*. 2012;48(2):1–36. doi: [10.18637/jss.v048.i02](https://doi.org/10.18637/jss.v048.i02).
 36. Csardi G, Nepusz T. The igraph software package for complex network research. *Int J Complex Syst*. 2005;1695.
 37. Ho D, Imai K, King G, Stuart E. MatchIt: nonparametric preprocessing for parametric causal inference. *J Stat Soft*. 2011;42(8):1–28. doi: [10.18637/jss.v042.i08](https://doi.org/10.18637/jss.v042.i08).
 38. Liddle PF, Ngan ETC, Duffield G, Kho K, Warren AJ. Signs and Symptoms of Psychotic Illness (SSPI): a rating scale. *Br J Psychiatry*. 2002;180(1):45–50. doi: [10.1192/bjp.180.1.45](https://doi.org/10.1192/bjp.180.1.45).
 39. Peralta V, Gil-Berrozpe GJ, Librero J, Sánchez-Torres A, Cuesta MJ. The symptom and domain structure of psychotic disorders: a network analysis approach. *Schizophr Bull*. 2020;1(1):sgaa008. doi: [10.1093/schizbullopen/sgaa008](https://doi.org/10.1093/schizbullopen/sgaa008).
 40. Liddle PF, Ngan ETC, Caissie SL, et al. Thought and Language Index: an instrument for assessing thought and language in schizophrenia. *Br J Psychiatry*. 2002;181(04):326–330.
 41. Korta K, Perry J. Pragmatics. In: *The Stanford Encyclopedia of Philosophy*. Palo Alto, California: Stanford University; 2019. <https://plato.stanford.edu/archives/spr2020/entries/pragmatics/>.
 42. Grice HP. Utterer's meaning, sentence-meaning, and word-meaning. In: Kulas J, Fetzer JH, Rankin TL, eds. *Philosophy, Language, and Artificial Intelligence*. Vol 2. Studies in Cognitive Systems, Springer Netherlands; 1968:49–66. doi: [10.1007/978-94-009-2727-8_2](https://doi.org/10.1007/978-94-009-2727-8_2).
 43. Grosz BJ, Sidner CL. Attention, intentions, and the structure of discourse. *Comput Linguist*. 1986;12(3):30.
 44. Grosz BJ, Joshi AK, Weinstein S. *Centering: A Framework for Modelling the Local Coherence of Discourse*. Philadelphia, Pennsylvania: Defense Technical Information Center; 1995. doi: [10.21236/ADA324949](https://doi.org/10.21236/ADA324949).
 45. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language use: our words, our selves. *Annu Rev Psychol*. 2003;54(1):547–577. doi: [10.1146/annurev.psych.54.101601.145041](https://doi.org/10.1146/annurev.psych.54.101601.145041).
 46. Verhoeven L, Vermeer A. Communicative competence and personality dimensions in first and second language learners. *Appl Psycholinguist*. 2002;23(3):361–374. doi: [10.1017/S014271640200303X](https://doi.org/10.1017/S014271640200303X).
 47. Koops S, Brederoo SG, de Boer JN, Nadema FG, Voppel AE, Sommer IE. Speech as a biomarker for depression. 2021. doi: [10.2174/1871527320666211213125847](https://doi.org/10.2174/1871527320666211213125847).
 48. Cohen AS, Alpert M, Nienow TM, Dinzeo TJ, Docherty NM. Computerized measurement of negative symptoms in schizophrenia. *J Psychiatr Res*. 2008;42(10):827–836. doi: [10.1016/j.jpsychires.2007.08.008](https://doi.org/10.1016/j.jpsychires.2007.08.008).
 49. Parola A, Simonsen A, Bliksted V, Fusaroli R. Voice patterns in schizophrenia: a systematic review and Bayesian meta-analysis. *Schizophr Res*. 2020;216:24–40. doi: [10.1016/j.schres.2019.11.031](https://doi.org/10.1016/j.schres.2019.11.031).