Schizophrenia Bulletin Open
doi:10.1093/schizbullopen/sgab010

Small Words That Matter: Linguistic Style and Conceptual Disorganization in Untreated First-Episode Schizophrenia

Angelica Silva*[,1], Roberto Limongi[1,2], Michael MacKinley[1–3], and Lena Palaniyappan[1–5]

[1]Robarts Research Institute, London, ON, Canada; [2]Department of Psychiatry, University of Western Ontario, London, ON, Canada; [3]Lawson Health Research Institute, London, ON, Canada; [4]Department of Medical Biophysics, University of Western Ontario, London, ON, Canada; [5]The Brain and Mind Institute, University of Western Ontario, London, ON, Canada

*To whom correspondence should be addressed; Robarts Research Institute, 1151 Richmond Street N, London, ON, Canada N6A 5B7; fax: 5199315789, e-mail: asilva43@uwo.ca

This study aimed to shed light on the linguistic style affecting the communication discourse in first-episode schizophrenia (FES) by investigating the analytic thinking index in relation to clinical scores of conceptual and thought disorganization (Positive and Negative Syndrome Scale, PANSS-P2 and Thought and Language Index, TLI). Using robust Bayesian modeling, we report three major findings: (1) FES subjects showed reduced analytic thinking, exhibiting a less categorical linguistic style than healthy control (HC) subjects (Bayes factor, BF_{10} > 1000), despite using the same proportion of function and content words as HCs; (2) the lower the analytic thinking score, the higher the symptoms scores of conceptual disorganization (PANSS-P2, BF = 22.66) and global disorganization of thinking (TLI, BF_{10} = 112.73); (3) the linguistic style is a better predictor of conceptual disorganization than the cognitive measure of processing speed in schizophrenia (SZ). These findings provide an objectively detectable linguistic style with a focus on Natural Language Processing Analytics of transcribed speech samples of patients with SZ that require no clinical judgment. These findings also offer a crucial insight into the primacy of linguistic structural disruption in clinically ascertained disorganized thinking in SZ. Our work contributes to an emerging body of literature on the psychopathology of SZ using a first-order lexeme-level analysis and a hypothesis-driven approach. At a utilitarian level, this has implications for improving educational and social outcomes in patients with SZ.

Key words: first-episode schizophrenia/formal thought disorder/speech–language impairment/natural language processing analytics/linguistic style

Introduction

Human language is a multi-level system strongly associated with thought organization[1,2] and many aspects of neurocognition (e.g. executive functioning).[3] The form and content of one's linguistic expression continues to be the main substrate of clinical diagnosis for psychotic disorders. A comprehensive understanding of "schizophrenic language"—a term coined by Chaika[4] as a special case of linguistic disruption—requires the study of clinical deviance[5,6] as well as variations in normal speech.[7] Linguistic approaches offer specific conceptual and procedural frameworks through Natural Language Processing analytics[8] to investigate the differential mechanisms[9] underlying a wide-ranging set of communication-discourse disturbances.[10] In this work, we focus on the relationship between these disruptions in transcribed speech samples and the clinical impression of formal thought disorder (FTD) in schizophrenia (SZ). We directly address the need to use computational linguistics in a hypothesis-driven approach, to move from exploratory to explanatory and predictive frameworks.[9]

Most previous studies on language in SZ fall in two approaches. First, a symptom-primacy approach that describes deviations in speech/language with[7,11] or without[12,13] linguistic concepts. The approach generally focusses on the pathological consequences of communication disturbances[14,15] or disordered thoughts[16,17] and relies on cross-sectional interviews. Secondly, a linguistic-primacy approach considers speech disturbance in SZ to be a variation in the normal language use. This approach focusses on studying the quantity and quality of words usage, quantifying speech incoherence and textual cohesion,[18,19] and detecting failures in building meaning including converting thoughts into expressions (c.f. referentiality).[20–22] Irrespective of the differences in the perspectives, three points of consensus emerge from these approaches. First, language can be affected by cognitive and thought disorders in any of the three different stages of production (i.e. conceptualization, formulation, and articulation).[23] Secondly, speech/language disturbances have been observed not only in SZ but also in other disorders, in apparently healthy individuals, and individuals at risk of psychosis.[24,25] Thirdly, in most patients, speech/language disturbances are subtle and thus elusive for the listener, but when severe, they reveal an unexpected linguistic pattern that stands out (likely creating the verbal equivalent of Rumke's "praecox feeling"[26]). Despite this consensus, to date, the study of failures in spoken language production in acute SZ continues to be elusive—especially because language sampling outside clinical settings is particularly challenging.

Failures in spoken language production can be captured through an analysis of not only its content (what is said) but also its form (how it is said).[23] The analysis of *linguistic styles* is an emergent computational linguistic approach to assess the form of language, i.e. "how people put their words together to create a message."[27] One style is *categorical*, suggesting formal and hierarchical thinking patterns. The other style is *narrative*, suggesting more intuitive and episodic patterns. A singular dimension capturing these two opposing poles can be extracted using computer-based analysis at the most basic first-order lexeme level of language. This dimension is the *analytic thinking index* whose numerical values (in a 0–100 scale) reflect a bipolar narrative-categorical linguistic style continuum.[28–31] This index is based on the proportional use of eight function words (FW) (articles, prepositions, conjunctions, nonreferential adverbs, negations, and auxiliary verbs). As elaborated upon (see Data Analysis Procedure), articles and prepositions take positive values—driving the linguistic style toward the continuum's categorical pole. Conversely, all the other FWs take negative values which drive the linguistic style toward the narrative pole. The positive or negative value of FW stems from a principal component analysis performed on a large corpus comprised by 50 000 essays written by first-year university students.[32]

One aspect of the analytic thinking index requires consideration: pronouns and articles showed opposite loadings. From a syntax-centered linguistics perspective, both these FWs are the type of words needed to express referentiality and thus expected to be grouped together. Nevertheless, in corpus linguistics, pronouns (unlike articles) endorse the text's focus—a function that emerges from interpreting co-occurrence linguistic patterns—with predominant narrative style. See, for example, the seminal work of Biber.[33] Therefore, pronouns drive the analytic thinking score closer to the narrative pole (wherein dynamic, personal narratives are dominant), whereas articles drive the score closer to the categorical pole (whereby complex concepts/objects are organized).

By relying on the statistical co-occurrence of FWs in language use, the analytic thinking score provides an intuitive yet robust index to assess the complexity of the conceptual organization reflected in speech/language samples. A low score in analytic thinking (driven by, for example, a low proportion of articles and a high proportion of conjunctions) would indicate a less categorical style, whereas a higher score would reflect a heightened abstract thinking and syntactic complexity (greater use of articles and prepositions). Subjects with higher analytic thinking scores are likely to deconstruct complex concepts into more manageable parts and better present their interrelationships. In what follows, we apply the concept of analytic thinking as a marker of the linguistic style in SZ. We investigate whether this linguistic style directly speaks to how patients organize their thoughts (c.f. conceptual disorganization) and express them.

It is a longstanding tenet in cognitive science that highly organized conceptual knowledge equals high analytic thinking.[34] From this theoretical assumption, it follows that conceptual disorganization would imply lowly organized conceptual knowledge. Therefore, we predicted lower analytic thinking (i.e. less conceptually organized) in FES patients than in healthy control (HC) subjects during a time-limited discourse referring to an unambiguous stimulus (pictures) [hypothesis 1]. Furthermore, we expected the reduced analytic thinking index to relate to higher severity of clinician-measured indices of conceptual disorganization (PANSS and TLI scales) [hypothesis 2]. Previous works have focussed on linking disorganization to neurocognition rather than to linguistic structure.[35] In particular, Ventura et al. demonstrated that reduced processing speed, one of the most consistent cognitive deficits observed in patients with SZ,[36–39] was the best neurocognitive predictor of clinically measured disorganization.[40] We tested whether linguistic style could explain disorganization better than processing speed. As antipsychotic exposure modifies the degree of disorganization, we recruited a challenging sample of first-episode patients with acute untreated episode of psychosis (<3 days of median antipsychotic dose exposure) with longitudinal follow-up to confirm diagnostic stability. This study is part of a pre-registered observational trial whose primary aim is to characterize the thought and language disorder in SZ (NCT02882204).

Method

Participants

The sample size was 78 English-speaking subjects, divided in 48 treatment-naive FES subjects (38 males) of mean (SD) age = 21.89 (3.63) years and 30 HC subjects (18 males) of mean (SD) age = 21.62 (3.47) years. The clinical sample was obtained from consecutive new referrals to the PEPP (Prevention and Early Intervention for Psychosis Program) at London Health Sciences Centre, London, Ontario, Canada between April 2017 and July 2019. Patients were in the acute phase of illness and recruited upon referral, irrespective of hospitalization status, before antipsychotic treatment was established. The PEPP uses an assertive case-management model to provide assessment and treatment to individuals 16–39 years old experiencing FES. All potential study participants provided written informed consent prior to participation as per the approval provided by the Western University Health Sciences Research Ethics Board, London, Ontario, Canada. All participants received a consensus diagnosis from a minimum of 3 psychiatrists (2 research psychiatrists and the primary treatment provider from the PEPP clinic) after approximately 6 months on the basis of the best estimate procedure (as described in Leckman et al[41] and the Structured Clinical Interview for DSM-5[42]). The 6-month consensus diagnosis confirmed that none of the included participants met criteria for bipolar disorder with psychotic features, major depressive disorder with psychotic features, or drug-induced psychoses.

HC subjects were recruited from the community through posters, had no personal history of mental illness, and no family history of psychotic disorders. Group matching with the FES cohort for age, sex, and parental socioeconomic status was maintained. Exclusion criteria for both groups involved meeting criteria for a substance-use disorder in the past year according to DSM-5 criteria,[42] having a history of a major head injury (leading to a significant period of unconsciousness or seizures), having a significant, uncontrolled medical illness, the presence of intellectual/developmental disorder, >2 weeks of lifetime antipsychotic exposure, or the inability to provide informed consent. We constructed an (ordinal) education level variable using the Statistics Canada classification of highest educational attainment[43] (Supplementary table). Complete demographic and clinical data are provided in table 1.

Instruments

Psychiatric Symptoms and Cognitive Assessment. The Positive and Negative Syndrome Scale-8 Item (PANSS-8) is a condensed version of the interview-based measure PANSS for psychosis, applied by one of the 2 research psychiatrists [intraclass correlation (ICC); 2 raters; 10 subjects =0.91]. To quantify processing speed, we used a modified digit symbol substitution task (DSST).[44] The written and oral items of the DSST were scored separately, and the items' mean scores were used for assessment as in prior studies.[45,46] Processing speed score

Table 1. Demographic, Clinical, Cognitive Scores

	HC	FES	BF10	95% HDI
Demographic and Clinical Data				
Age (Years)	21.89 (3.24)	21.43 (3.24)	0.29	−0.56, 0.32
Sex (Male)	38%	62%		
Education Scale	2.86 (1.06)	2.30 (1.02)	2.31	−0.94, −0.03
SES	3.4 (1.25)	3.24 (1.43)	0.273	−0.23, 0.54
PANSS (Total)		26.38 (6.62)		
PANSS-P2		3.26 (1.66)		
TLI (Total)		1.17 (1.2)		
Impoverishment of Thinking		0.44 (0.65)		
Disorganization of Thinking		0.73 (0.94)		
DUP (Months)		8.82 (13.17)		
DDD-Days		2.09 (3.53)		
SOFAS	82.15 (4.48)	38.38 (12.65)	>1000	−49.97, −36.94
DSST	68.28 (10.95)	52.79 (13.33)	>1000	−1.67, −0.66
Linguistic Data				
Words per Sentence	13.82 (3.23)	11.56 (2.52)	147.5	−1.33, 0.37
Articles	11.14 (2.93)	9.53 (3.57)	1.41	−0.87, 0.02
Prepositions	13.86 (2.72)	12.68 (3.18)	0.599	−0.73, 0.13
Personal Pronouns	6.61 (3.47)	7.94 (4.37)	0.931	−0.07, 0.81
Impersonal Pronouns	5.61 (2.38)	6.47 (3.05)	0.743	−0.10, 0.77
Auxiliary Verbs	10.44 (2.27)	11.32 (2.93)	1.4	−0.02, 0.87
Adverbs	6.47 (2.32)	7.02 (2.84)	0.525	−0.15, 0.71
Conjunctions	5.86 (2.22)	7.01 (3.21)	0.739	−0.10, 0.77
Negations	0.93 (2.03)	1.2 (1.28)	0.259	−0.39, 0.51

Note: FWs are presented in mean percentages relative to the total WC. Summary statistics mean (standard deviation). For completeness, Bayesian *t*-tests of between-group differences are shown. SES, socioeconomic status; PANSS, Positive and Negative Syndrome Scale; TLI, thought language index; DUP, duration of untreated psychosis; DDD, defined daily dose equivalents of antipsychotic medication; SOFAS, Social and Occupational Functioning Assessment Score; DSST, Digit Symbol Substitution Test.

refers to the number of symbols coded using matched digits within a specified time period (90 s); this test per se did not involve any word lists. These measures were obtained on the same week during the first clinical contact with symptoms of psychosis, prior to the patients receiving clinically adequate treatment. Using WHO's algorithm for defined daily doses (DDDs) for antipsychotic medications,[47] we derived a common unit of exposure to antipsychotics to quantify the baseline exposure. We also used the modified Social and Occupational Functioning Assessment Scale (SOFAS; administered by a single rater) to assess the overall level of functioning at the time of presentation.[48] We did not administer the longer version of PANSS or other detailed cognitive tests given the acute illness phase during which the data were gathered (Supplementary Material).

The Thought Language Index (TLI). The TLI[12] is an interview-based instrument to assess FTD under standardized conditions. We used a picture-speech task that induced participants to elaborate 1-min spontaneous speech (oral soliloquies) in response to three photographs from the Thematic Apperception Test[49] 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."

Researchers that administered the picture-speech task were blinded to participant status.[50] Responses were recorded, transcribed, and scored on eight domains which were integrated in two merged labels: (1) Impoverishment in Thinking which included poverty of speech, weakening of goal, preservation of ideas (ICC between 2 raters for 78 subjects = 0.97) and (2) Disorganization in Thinking which comprised looseness, peculiar use of words, peculiar sentences, peculiar logic, and distractibility (ICC=0.99). Finally, global scores for each of the merged labels accomplished the TLI (i.e. global impoverishment in thinking and global disorganization in thinking). Cognitive assessments, the TLI interview, and rating were completed by trained graduate-level research assistants. Symptom assessments were performed in the clinical context, with clinical raters being blind to linguistic and cognitive scores.

Linguistic Inquiry Word Count (LIWC). Linguistic Inquiry Word Count Software (LIWC 2015 Edition) is a computational-lexical approach, essentially a word detector providing summaries of psycholinguistic dimensions (i.e. analytic thinking score) and pre-defined content word themes (e.g. negative emotion words) derived from psychometric rates. In the two-step process, LIWC analyzes the current target word contained in texts comparing and matching every single word against master dictionaries using its own language corpora composed of "almost 6,400 words, word steams, and selected emoticons from a sample of ~181,000 text files."[51] Secondly, a standard LIWC computes the percentage of co-occurrences. LIWC has recently gained attention in several research areas establishing the relationship between linguistic-thinking styles and both personality traits, see, for example,[31] and mental health conditions (Supplementary Material).

Data Analysis Procedure

The data analysis was completed in four steps. First, research assistants transcribed speech samples blinded to participant status. Secondly, a preprocessing was needed in order to use the LIWC software. Thirdly, one-by-one text files were entered into LIWC to obtain the standard output upon which subjects' analytic thinking scores were computed with this equation: Analytic Thinking = articles + prepositions – pronouns – auxiliary verbs – conjunctions – negations – nonreferential adverbs. The resulting algebraic sum was standardized in the 0–100 scale at the group level. Fourth, as detailed below, we performed two Bayesian analyses (using JASP, version 0.14) to evaluate our hypotheses.

Statistical (Bayesian) Analyses

First, we investigated whether the groups differed in the number of words (i.e. word count, WC) they used to describe the pictures. Furthermore, we investigated the effect of group in the percentage of FWs and CWs. Sex and education were included as covariates. Furthermore, text number (corresponding to each of the three task pictures) and subjects were included as random effects. Finally, we tested the main hypothesis of the effect of group on the analytic thinking score. Secondly, we tested our second hypothesis via Bayesian Pearson's correlations between analytic thinking scores (averaged across the 3 transcripts) and clinical scores (i.e. PANSS-P2, Global Impoverishment of Thinking, and Global Disorganization of Thinking).

In both analyses, we report Bayes factors against the null model (BF_{10}), unless otherwise specified. Briefly, if $BF_{10} < 0$, we accepted the null hypothesis, whereas if $BF > 3$ we accepted the alternative hypothesis. As an index of effect size, we report the proportion of variance accounted for by the winning model (R^2). At a parameter level, we report the 95% highest density interval (HDI) of the posterior distribution of credible values as well as the proportion of this distribution that differed from 0 (posterior proportion, PP). Parameters were estimated using Markov Chain Monte Carlo procedures and were assessed for convergence (Supplementary Material).

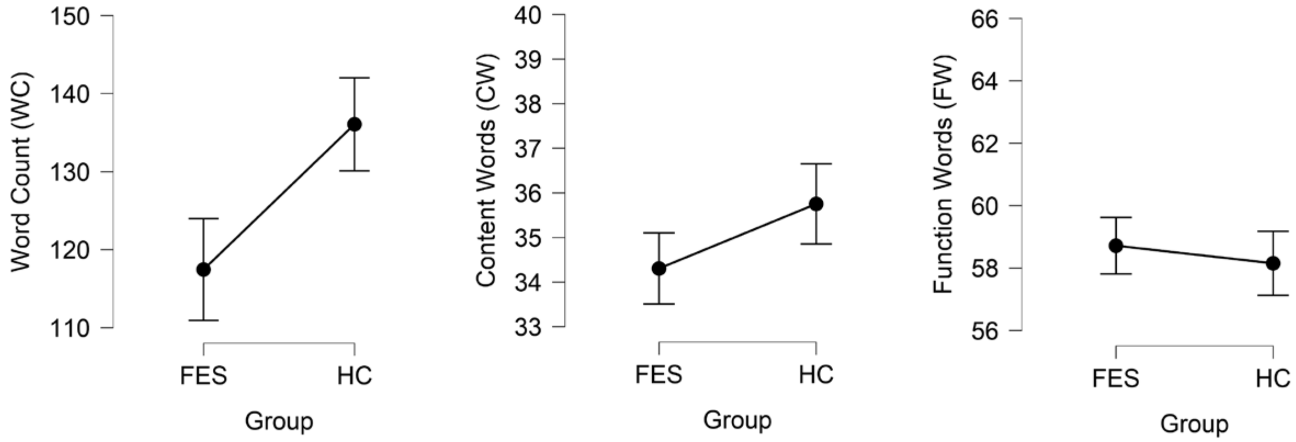

Fig. 1. Between-groups Bayesian comparison of total word count (WC), content words (CW), and function words (FWs). Error bars represent the 95% HDI.

Results

Group Effect on Analytic Thinking Score (Hypothesis 1)

FES subjects used lower number of words to describe the pictures than HCs (BF_{10} = 3.52; HC mean = 136.36, SD = 27.64; FES mean = 118.14, SD = 36.25; mode of the between-group difference = 17.8, PP = 1.0, 95% HDI [8.81, 26.5], R^2 = 0.77; figure 1). Moreover, both groups used the same percentage of FWs (BF_{10} = 0.31; HC mean = 58.1, SD = 4.76; FES mean = 58.67, SD = 5.04; mode of the between-group difference = 0.2, PP = 0.78, 95% HDI [−0.72, 1.72], R^2 = 0.5; figure 1). Similarly, both groups used the same percentage of CW (BF_{10} = 0.55; R^2 = 0.48; HC mean = 35.82, SD = 4.2; FES mean = 34.58, SD = 4.6; mode of the between-group difference = 1.15, PP = 0.97, 95% HDI [−0.01, 2.35]; figure 1). Note that in the case of CW, although the percentage of posterior credible values was 97%, the *most* credible of these values comprised negative estimates.

We confirmed our main hypothesis. The FES group showed a less categorical linguistic style than the HC group (mean-$_{FES}$ =54.09, SD = 30.42; mean-$_{HC}$ = 70.9, SD = 24.46; mode of the between-groups difference = −16.2, PP = 1.0, 95% HDI [−21.6, −11.1]; R^2 = 0.5; BF_{10} > 1000; figure 2). Table 2 shows that models including education level and sex received much less support from the data than the simpler model (see BF_M in table 2).

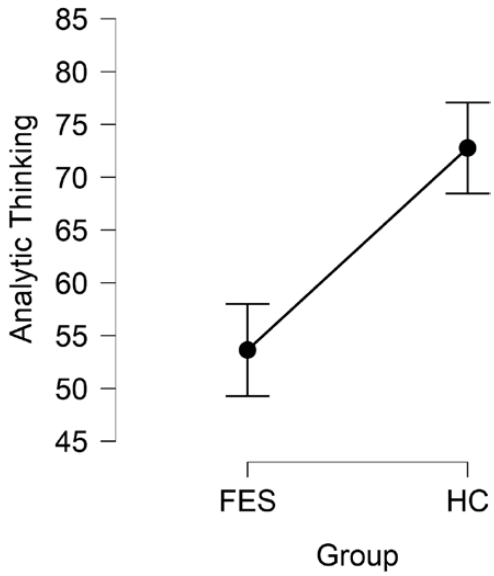

Fig. 2. Effects of group on the analytic thinking scores. Bars represent the 95% HDI.

Correlation Between Analytic Thinking and Clinical Symptoms (Hypothesis 2)

Analytic thinking was negatively correlated with both the PANSS-P2 score (BF_{10} = 22.66, r = −0.44, 95% HDI [−0.63, −0.17], PP = 0.99, figure 3) and the global disorganization of thinking score (BF_{10} = 112.73, r = −0.5, 95% HDI [−0.63, −0.17], PP = 1.0, figure 4). However, analytic thinking did not correlate with the global score of impoverishment of thinking (BF_{10} = 0.3, r = −0.15, 95% HDI [−0.4, 0.14], PP = 0.68, figure 5).

Given the compelling support that the data provided to analytic thinking as a symptoms predictor, this score emerges as a reliable index of conceptual disorganization. This suggests that analytic thinking could also be a better predictor than other indices of general cognitive functioning. To explore this potential generalization, we assessed whether analytic thinking is a better predictor than the DSST. When accounting for conceptual disorganization (PANSS-P2), analytic thinking outperformed the DSST (BF = 229). Furthermore, analytic thinking is at least 6 times as likely to account for global disorganization of thinking as is the DSST (BF = 6.9). For completeness, the DSST marginally outperformed analytic thinking when accounting for general impoverishment of

Table 2. Model Comparison

Models	$P(M)$	$P(M\|\text{data})$	BFM	BF10
Null Model	0.100	6.282e−7	5.654e−6	1.000
Group	0.100	0.502	9.058	798 400.521*
Group + Sex	0.100	0.277	3.456	441 678.968*
Group + Sex + Group × Sex	0.100	0.086	0.849	137 186.926*
Group + Education Scale	0.100	0.08	0.780	126 987.888*
Group + Sex + E. Scale	0.100	0.042	0.391	66 215.670*
Group + Sex + E. Scale + Group × Sex	0.100	0.013	0.122	21 253.776*
E. Scale	0.100	1.615e−7	1.454e−6	0.257
Sex	0.100	1.023e−7	9.210e−7	0.163
Sex + E. Scale	0.100	2.885e−8	2.596e−7	0.046

Note: All models included picture number as random effects. E. scale, education scale; $P(M)$, prior probability of rival models; $P(M|\text{data})$, probability of each model after seeing the data; BF_M, likelihood of each model compared with the mean of the rival models; BF_{10}, Bayes factor ratio of the model against the null. *The model outperformed the null. All complex models performed better than the null. However, no complex model performed better than the simpler model comprising only the effect of group.

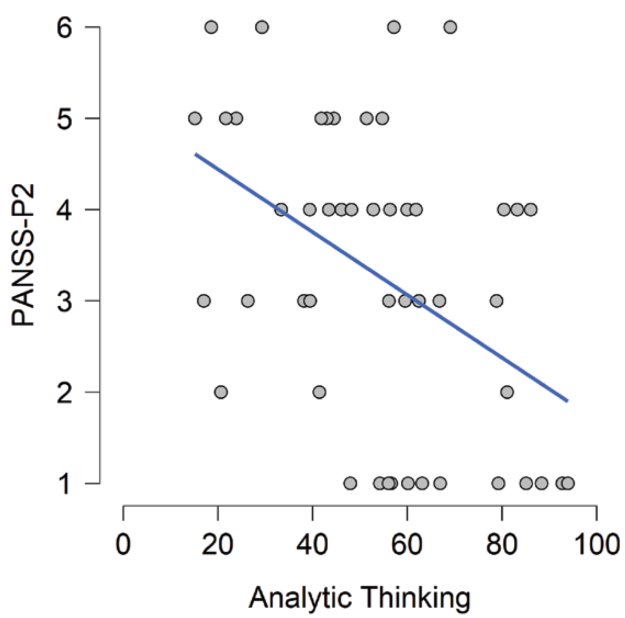

Fig. 3. Bayesian Pearson's correlation between analytic thinking and conceptual disorganization (PANSS-P2).

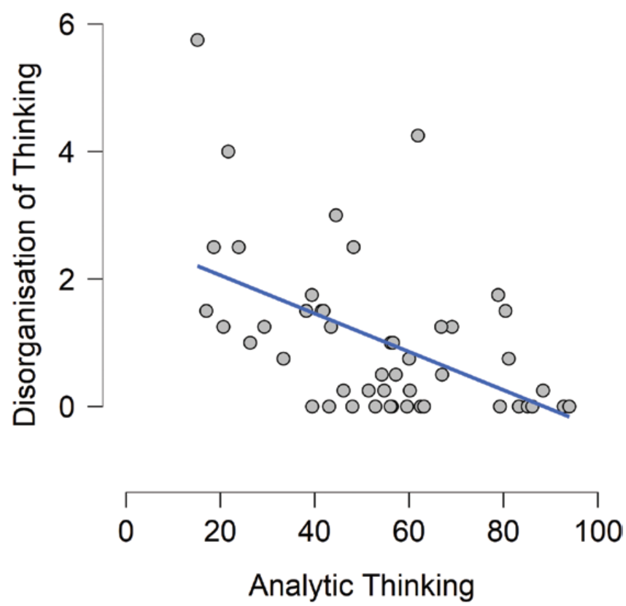

Fig. 4. Bayesian Pearson's correlation between analytic thinking and disorganization of thinking.

thinking (BF = 3.77). However, DSST barely won over the null model ($BF_{10} > 1.48$).

General Discussion

This study sheds light on the linguistic style affecting the communication discourse in FES by investigating the analytic thinking index in relation to clinical scores—PANSS-P2 and TLI. We report three major findings: (1) FES subjects showed reduced analytic thinking, exhibiting a less categorical linguistic style than HC subjects, despite using the same proportion of function and content words as HC; (2) the lower the analytic thinking score, the higher the syndrome scores of conceptual disorganization (PANSS-P2) and global disorganization of thinking (TLI); and (3) the linguistic style is a better predictor of conceptual disorganization than the cognitive measure of processing speed in SZ. These findings provide an objectively detectable linguistic style in SZ that requires no clinical judgment. Additionally, they provide a crucial insight into the primacy of linguistic structural disruption in the clinical impression of disorganized thinking in SZ.

We found support to the differential effect of group on analytic thinking as an emergent feature of the entire language structure. Remarkably, no single word type drove low analytic thinking scores in SZ (table 1). FEP and HC subjects were alike in the relative proportion of FW and CW usage. A deficit in FW acquisition could explain the lower analytic thinking score in the FES group. Children learn FW as combining words. Whereas CW (open-class) are often learned from single-word utterances matching

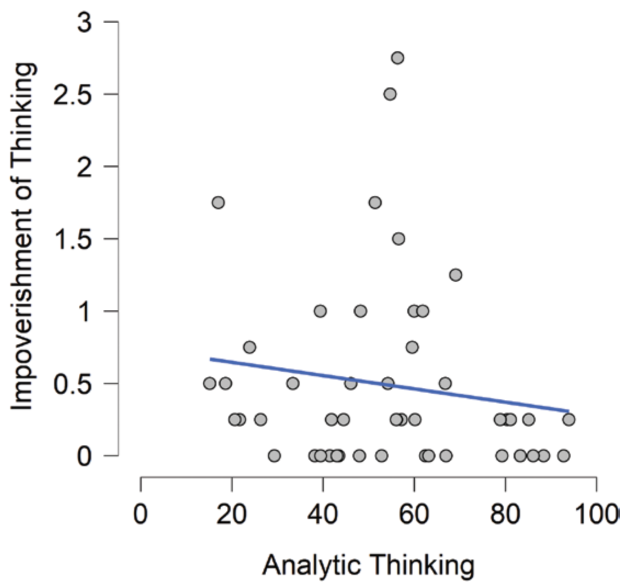

Fig. 5. Bayesian Pearson's correlation between analytic thinking and impoverishment of thinking.

word meanings to the concept-world knowledge, a multiword input is needed to learn FWs (closed-class). It also means that one condition for FW acquisition is syntactic proficiency.[52] Therefore, FWs play a dual role in helping syntax to maintain the relationship between words and sentence organization. Creating references needs a specific word-use profile that joints CW-FW support meaning at the semantic level. In particular, the appropriate selection, rather than the learning of FW itself, may be disrupted in SZ.

We observed a greater role for the linguistic style, rather than a domain-general neurocognitive score (processing speed), in explaining the variance of disorganized thinking. In contrast, Bora et al[35] highlighted a major role for neurocognition (working memory, processing speed deficits) as well as syntactic comprehension deficits in disorganization in SZ (more specifically, positive FTD). While it is tempting to rekindle the debate on linguistic vs. neurocognitive primacy in speech disorder of SZ, it is important to note that we were limited in our ability to extensively measure cognition during the acute psychotic phase. Attention, working memory, and source monitoring indices may prove to be better predictors than the DSST in future studies.

Speaking results from online processing in which *errors* in spoken language are common and expected. However, normal speakers correct themselves after noticing an error during their speech.[53] One major problem in SZ patients is that they do not correct messages[7] which speaks to cognitive control impairment.[54] Docherty et al[55] related the communication failures in disorganized speech to sequencing impairment.[56] Chaika[57] argued that SZ subjects also deliver verbal messages with less voluntary control. In the absence of an adequate control required for categorical partitions and sequencing, an effortless compensatory approach may rely more on an intuitive, less categorical, or abstract style of communication. As a result, the discourse stimulus (e.g. from a picture) becomes detail-oriented, replete with irrelevant pieces of information, less cohesive, and full of derailments instead of being impoverished in communicative function (Supplementary Material).

Compensatory phenomena in SZ are present at structural,[58] physiological, and cognitive levels.[59] We speculate that the apparent grammatical incompetence and resulting referential anomalies may be an unintended ploy to maintain discourse in the presence of cognitive control deficits. Neurophysiological studies capturing the role of cognitive control networks during discursive tasks, e.g. using effective connectivity models of brain function (e.g. fMRI),[60] will be able to clarify this speculation.

Analytic thinking accounted for 19% of the variance of the PANSS-P2 scores and for 25% of the variance of the TLI disorganized thinking scores. In both cases, a large proportion of variance remained unexplained. In general, the analytic thinking score is a behavioral index. Our previous work has suggested that parameter estimates of computational models of behavioral measures are better predictors of clinical scores than the behavioral measures per se, and this predictive power is augmented when they are combined with neurochemical, computational, and network connectivity parameters in a single model.[59]

In closing, our observations have key implications for understanding and treating SZ. First, they support a compelling body of literature on the linguistic basis of the psychopathology of SZ, recently revived by Hinzen et al[61] in the pursuit of delusion formation or the hallucinated voice talk by Tovar et al.[62] More broadly, the reduced deployment of referential function in those with a less categorical linguistic style may drive deficits in structural learning.[63,64] Secondly, they suggest a role of categorical thinking in learning processes[34] and its relevance to semantic long-term memory[65]—the permanent storage of organized knowledge. From this perspective, the less categorical style of language deployment may have implications for educational outcomes in SZ. In fact, academic success seems to require a more categorical linguistic style[32] due to the conventions of spoken and written academic discourses.[66] In this context, educational approaches to improve analytic thinking style may be beneficial for patients with SZ.

This work has several strengths including the recruitment of minimally medicated patients during an acute stage of psychosis, something that has not been achieved to date in prior linguistic studies of psychosis. Also several limitations must be considered. First, while LIWC allows us to quantify frequency occurrence of words, this does not account for contextual text meanings. Secondly, contextual, prosodic, and phonetic aspects of spoken language were not considered in this study. However, clinical employment of discourse analysis is more realistic with transcribed speech samples that are readily available with minimal resource constraints, allowing large-scale studies as well as future repeated measurements. Thirdly, the stimuli we used were not designed specifically to study thinking styles. While we employed a descriptive discourse schema,[67] emotional narrative tasks could possibly increase the frequency of type of FWs.[68] Such a descriptive task will lead the speaker toward talking about whole, parts, attributes, spatial location, comparisons,[67] and reflecting everyday communication demands crucial for the study of psychosis. Fourth, we lacked information about bilingualism and native language of our subjects, though English was the transactional language for all participants—no interpreters were used in any aspects of the study.

Supplementary Material

Supplementary data are available at *Schizophrenia Bulletin Open* online.

Funding

This study was funded by The Canadian Institutes of Health Research (CIHR) Foundation Grant (375104/2017) to L. P.; the Academic Medical Organization of Southwestern Ontario Opportunities Fund to L. P.; Bucke Family Fund to L. P.; and support from the Chrysalis Foundation to L. P.

Acknowledgments

We thank Dr Kara Dempster for recruitment and diagnostic assessments and Ms Hanna Ke, London, Ontario and Ms Karishma Singh, Ireland for transcription support. We are grateful to all patients and families who took part in this study. Our special thanks to PEPP physicians and staff members and Dr Ross Norman for kind assistance with training sessions for assessments. L. P. receives book royalties from Oxford University Press and income from the SPMM MRCPsych course. In the last 5 years, his or his spousal pension funds held shares of Shire Inc. and GlaxoSmithKline. L. P. has received investigator initiated educational grants from Otsuka Canada and Janssen Canada in 2017 and a speaker fee from Otsuka Canada in 2017 and Canadian Psychiatric Association in 2018. L. P. and M. M. received support from Boehringer Ingelheim to attend an investigator meeting in 2017. All other authors report no potential conflicts of interest.

Data availability and data processing scripts statement

The data, preprocessing pipeline, and data analysis scripts that support the findings of this study are available from the corresponding author upon reasonable request.

Author contributions

A. S. provided the theorical framework and hypotheses, performed speech data processing, and took the lead in writing the first draft of the manuscript. R. L. led the Bayesian analytical design and supervised A. S. to undertake the statistical analysis. M. M. collected the data and assisted in the analysis. L. P. sourced funding, assessed patients, designed and coordinated the project, supervised A. S. and M. M. All authors provided critical feedback and helped shape the interpretation and approved the final version of this manuscript.

References

1. Chomsky N. *The Minimalist Program*. Cambridge, MA: MIT Press; 1995.
2. Jackendoff R. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. New York: Oxford University Press; 2002.
3. Cardella V. *Language and Schizophrenia*. 1st ed. London: Routledge; 2018.
4. Chaika E. A linguist looks at "schizophrenic" language. *Brain Lang.* 1974;1(3):257–276.
5. Andreasen NC. Thought, language, and communication disorders. I. Clinical assessment, definition of terms, and evaluation of their reliability. *Arch Gen Psychiatry.* 1979;36(12):1315–1321.
6. Andreasen NC, Grove WM. Thought, language, and communication in schizophrenia: diagnosis and prognosis. *Schizophr Bull*. 1986;12(3):348–359.
7. Covington MA, He C, Brown C, *et al.* Schizophrenia and the structure of language: the linguist's view. *Schizophr Res*. 2005;77(1):85–98.
8. Corcoran CM, Cecchi GA. Using language processing and speech analysis for the identification of psychosis and other disorders. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2020;5(8):770–779.
9. Hitczenko K, Mittal VA, Goldrick M. Understanding language abnormalities and associated clinical markers in psychosis: the promise of computational methods. *Schizophr Bull.* 2020;47(3):1–19.
10. Corcoran C, Mittal V, Bearden C, *et al.* Language as a biomarker for psychosis: A natural language processing approach. *Schizophr Res*. 2020;226:158–166.
11. Ceccherini-Nelli A, Crow TJ. Disintegration of the components of language as the path to a revision of Bleuler's and Schneider's concepts of schizophrenia. Linguistic disturbances compared with first-rank symptoms in acute psychosis. *Br J Psychiatry*. 2003;182:233–240.
12. Liddle PF, Ngan ET, Caissie SL, *et al.* Thought and language index: an instrument for assessing thought and language in schizophrenia. *Br J Psychiatry*. 2002;181:326–330.
13. 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.
14. Docherty NM, DeRosa M, Andreasen NC. Communication disturbances in schizophrenia and mania. *Arch Gen Psychiatry*. 1996;53(4):358–364.
15. Bambini V, Arcara G, Bechi M, Buonocore M, Cavallaro R, Bosia M. The communicative impairment as a core feature of

schizophrenia: Frequency of pragmatic deficit, cognitive substrates, and relation with quality of life. *Compr Psychiatry*. 2016;71:106–120.
16. Johnston MH, Holzman PS, Solovay MR, *et al.* Scoring manual for the Thought Disorder Index. *Schizophr Bull*. 1986;12(3):483–496.
17. Hinzen W, Rosselló J. The linguistics of schizophrenia: Thought disturbance as language pathology across positive symptoms. *Front Psychol*. 2015;6:971.
18. MacKinley M, Chan J, Ke H, Dempster K, Palaniyappan L. Linguistic determinants of formal thought disorder in first episode psychosis. *Early Interv Psychiatry*. 2021;15(2):344–351.
19. Elvevåg B, Foltz PW, Weinberger DR, Goldberg TE. Quantifying incoherence in speech: an automated methodology and novel application to schizophrenia. *Schizophr Res*. 2007;93(1–3):304–316.
20. Kuperberg GR. Building meaning in schizophrenia. *Clin EEG Neurosci*. 2008;39(2):99–102.
21. Mota NB, Vasconcelos NA, Lemos N, *et al.* Speech graphs provide a quantitative measure of thought disorder in psychosis. *PLoS One*. 2012;7(4):e34928.
22. Çokal D, Sevilla G, Jones WS, *et al.* The language profile of formal thought disorder. *NPJ Schizophr*. 2018;4(1):18.
23. Voleti R, Liss JM, Berisha V. A review of automated speech and language features for assessment of cognitive and thought disorders. *IEEE J Select Top Signal Process*. 2020;14(2):282–298.
24. 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.
25. Gooding DC, Ott SL, Roberts SA, Erlenmeyer-Kimling L. Thought disorder in mid-childhood as a predictor of adulthood diagnostic outcome: findings from the New York High-Risk Project. *Psychol Med*. 2013;43(5):1003–1012.
26. Pallagrosi M, Fonzi L. On the concept of praecox feeling. *Psychopathology*. 2018;51(6):353–361.
27. Pennebaker JW, Chung C. The psychological functions of functions words. In: Fiedler K, ed. *Social Communication*. NY: Psychological Press; 2007:343–359.
28. McHaney R, Tako A, Robinson S. Using LIWC to choose simulation approaches: a feasibility study. *Decis Supp Syst.* 2018;111:1–12.
29. Pennebaker JW, King L. Linguistic styles: language use as an individual difference. *J Person Social Psychol.* 2000;77:1296–1312.
30. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language use: our words, our selves. *Annu Rev Psychol*. 2003;54:547–577.
31. Jordan KN, Sterling J, Pennebaker JW, Boyd RL. Examining long-term trends in politics and culture through language of political leaders and cultural institutions. *Proc Natl Acad Sci USA*. 2019;116(9):3476–3481.
32. Pennebaker JW, Chung CK, Frazee J, Lavergne GM, Beaver DI. When small words foretell academic success: the case of college admissions essays. *PLoS One*. 2014;9(12):e115844.
33. Biber D. *Variation Across Speech and Writing*. Cambridge, UK: Cambridge University Press; 1988.
34. Bruner JS, Goodnow JJ, Austin GA. *A Study of Thinking*. New York: Wiley; 1956.
35. Bora E, Yalincetin B, Akdede BB, Alptekin K. Neurocognitive and linguistic correlates of positive and negative formal thought disorder: A meta-analysis. *Schizophr Res*. 2019;209:2–11.
36. Nuechterlein KH, Subotnik KL, Green MF, *et al.* Neurocognitive predictors of work outcome in recent-onset schizophrenia. *Schizophr Bull*. 2011;37(suppl 2):S33–S40.
37. Mohamed S, Paulsen JS, O'Leary D, Arndt S, Andreasen N. Generalized cognitive deficits in schizophrenia: a study of first-episode patients. *Arch Gen Psychiatry*. 1999;56(8):749–754.
38. Dickinson D. Digit symbol coding and general cognitive ability in schizophrenia: worth another look? *Br J Psychiatry*. 2008;193(5):354–356.
39. Dickinson D, Ramsey ME, Gold JM. Overlooking the obvious: a meta-analytic comparison of digit symbol coding tasks and other cognitive measures in schizophrenia. *Arch Gen Psychiatry*. 2007;64(5):532–542.
40. Ventura J, Thames AD, Wood RC, Guzik LH, Hellemann GS. Disorganization and reality distortion in schizophrenia: a meta-analysis of the relationship between positive symptoms and neurocognitive deficits. *Schizophr Res*. 2010;121(1–3):1–14.
41. Leckman JF, Sholomskas D, Thompson WD, Belanger A, Weissman MM. Best estimate of lifetime psychiatric diagnosis: a methodological study. *Arch Gen Psychiatry*. 1982;39(8):879–883.
42. APA. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. Arlington, VA: American Psychiatric Publishing, Inc.; 2013.
43. Canada S. Education in Canada: Attainment, Field of Study and Location of Study. Available at: https://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-012-x/99-012-x2011001-eng.cfm#a2.
44. Palaniyappan L, Al-Radaideh A, Mougin O, Gowland P, Liddle PF. Combined white matter imaging suggests myelination defects in visual processing regions in schizophrenia. *Neuropsychopharmacology*. 2013;38(9):1808–1815.
45. Palaniyappan L, Simmonite M, White TP, Liddle EB, Liddle PF. Neural primacy of the salience processing system in schizophrenia. *Neuron*. 2013;79(4):814–828.
46. Briley PM, Liddle EB, Simmonite M, *et al.* Regional brain correlates of beta bursts in health and psychosis: a concurrent electroencephalography and functional magnetic resonance imaging study. *Biol Psychiatr: Cogn Neurosci Neuroimaging*. 2020.
47. WHO. Guidelines for ATC Classification and DDD Assignment; 2013.
48. Morosini PL, Magliano L, Brambilla L, Ugolini S, Pioli R. 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*. 2000;101(4):323–329.
49. Murray HA. Thematic Apperception Test Manual, by Henry A. Murray, M.D., and the Staff of the Harvard Psychological Clinic. Cambridge, MA: Harvard University Press; 1943.
50. Sommer IE, Derwort AM, Daalman K, de Weijer AD, Liddle PF, Boks MP. Formal thought disorder in non-clinical individuals with auditory verbal hallucinations. *Schizophr Res*. 2010;118(1–3):140–145.
51. Pennebaker JW, Boyd R, Jordan K, Blackburn K. *The Development and Psychometric Properties of LIWC2015*. Austin, TX: University of Texas at Austin; 2015.

52. Ninio A. Learning to structure sentences through the acquisition of grammatical words: introduction to the special issue on the role of grammatical words in young children's syntactic development. *First Lang*. 2018;39(1):3–8.
53. Nozari N, Freund M, Breining B, Rapp B, Gordon B. Cognitive control during selection and repair in word production. *Lang Cogn Neurosci*. 2016;31(7):886–903.
54. Kuperberg GR. Language in schizophrenia. Part 1: an introduction. *Lang Linguist Compass*. 2010;4(8):576–589.
55. Docherty NM, Strauss ME, Dinzeo TJ, St-Hilaire A. The cognitive origins of specific types of schizophrenic speech disturbances. *Am J Psychiatry*. 2006;163(12):2111–2118.
56. Docherty NM, Hall MJ, Gordinier SW, Cutting LP. Conceptual sequencing and disordered speech in schizophrenia. *Schizophr Bull*. 2000;26(3):723–735.
57. Chaika E. *Understanding Psychotic Speech: Beyond Freud and Chomsky*. Springfield, IL, England: Charles C Thomas Publisher; 1990.
58. 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.
59. Limongi R, Jeon P, Mackinley M, *et al.* Glutamate and dysconnection in the salience network: neurochemical, effective connectivity, and computational evidence in schizophrenia. *Biol Psychiatry*. 2020;88(3):273–281.
60. Friston KJ, Preller KH, Mathys C, *et al.* Dynamic causal modelling revisited. *Neuroimage*. 2019;199:730–744.
61. Hinzen W, Rosselló J, McKenna P. Can delusions be understood linguistically? *Cogn Neuropsychiatry*. 2016;21(4):281–299.
62. Tovar A, Fuentes-Claramonte P, Soler-Vidal J, *et al.* The linguistic signature of hallucinated voice talk in schizophrenia. *Schizophr Res*. 2019;206:111–117.
63. Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND. How to grow a mind: statistics, structure, and abstraction. *Science*. 2011;331(6022):1279–1285.
64. Kemp C, Tenenbaum JB. The discovery of structural form. *Proc Natl Acad Sci USA*. 2008;105(31):10687–10692.
65. Tulving E, Craik FIK. *The Oxford Handbook of Memory*. New York, NW: Oxford University Press; 2000.
66. Hyland K. Academic discourse: English in a global context. NY: Continuum; 2009.
67. Buck B, Penn DL. Lexical characteristics of emotional narratives in schizophrenia: relationships with symptoms, functioning, and social cognition. *J Nerv Ment Dis*. 2015;203(9):702–708.
68. Hamon P, Baudoin P. Rhetorical status of the descriptive. *Yale French Studies* 1981;61:1–26.