Psychiatry Research 333 (2024) 115752

Contents lists available at ScienceDirect

Psychiatry Research

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

Navigating the semantic space: Unraveling the structure of meaning in psychosis using different computational language models

Rui He [a,*], Claudio Palominos [a], Han Zhang [a], Maria Francisca Alonso-Sánchez [b], Lena Palaniyappan [c,d,e], Wolfram Hinzen [a,f]

[a] Department of Translation & Language Sciences, Universitat Pompeu Fabra, Carrer Roc Boronat, 138, Barcelona, 08018, Spain
[b] CIDCL, Escuela de Fonoaudiología, Universidad de Valparaíso, Valparaíso, Chile
[c] Douglas Mental Health University Institute, Department of Psychiatry, McGill University, Montreal, Quebec, Canada
[d] Department of Medical Biophysics, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada
[e] Robarts Research Institute, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada
[f] Intitut Català de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

ARTICLE INFO

Keywords:
Connected speech
Incoherence
Semantic similarity
Semantic perplexity
Language model
Loosening of associations
Schizophrenia

ABSTRACT

Speech in psychosis has long been ascribed as involving 'loosening of associations'. We pursued the aim to elucidate its underlying cognitive mechanisms by analysing picture descriptions from 94 subjects (29 healthy controls, 18 participants at clinical high risk, 29 with first-episode psychosis, and 18 with chronic schizophrenia), using five language models with different computational architectures: FastText, which represents meaning non-contextually/statically; BERT, which represents contextual meaning sensitive to grammar and context; Infersent and SBERT, which provide sentential representations; and CLIP, which evaluates speech relative to a visual stimulus. These models were used to quantify semantic distances crossed between successive tokens/sentences, and semantic perplexity indicating unexpectedness in continuations. Results showed that, among patients, semantic similarity increased when measured with FastText, Infersent, and SBERT, while it decreased with CLIP and BERT. Higher perplexity was observed in first-episode psychosis. Static semantic measures were associated with clinically measured impoverishment of thought and referential semantic measures with disorganization. These patterns indicate a shrinking conceptual semantic space as represented by static language models, which co-occurs with a widening in the referential semantic space as represented by contextual models. This duality underlines the need to separate these two forms of meaning for understanding mechanisms involved in semantic change in psychosis.

List of acronyms

BERT	bidirectional encoder representations from transformers
BLIPS	brief and limited intermittent psychosis
CHR	clinical high-risk
CLIP	contrastive language-image pretraining
CS	chronic schizophrenia
FastText	Library for efficient text classification and representation learning
FEP	first-episode psychosis
FTD	formal thought disorder
fMRI	functional magnetic resonance imaging
GEE	Generalized estimating equation
HC	health control
LM	language model
MATTR	moving-averaged type-token-ratio
PANSS	positive and negative syndrome scale-8 items version
PPL	perplexity
PPPL	pseudo-perplexity
SIPS	brief structured interview for psychosis-risk syndromes
SOPS	scale of prodromal symptoms
SBERT	sentence BERT
TLI	thought language index

* Corresponding author.
E-mail address: rui.he@upf.edu (R. He).

https://doi.org/10.1016/j.psychres.2024.115752
Received 28 July 2023; Received in revised form 16 January 2024; Accepted 21 January 2024
Available online 23 January 2024

1. Introduction

Atypical forms of discourse in psychosis have long been described clinically as tangentiality or incoherence (Andreasen, 1979). These were conceptualized early on as revealing 'loosening of associations' (Bleuler, 1911). A critical desideratum is to understand the mechanisms involved. In this regard, speech production inherently involves the integration of two very different kinds of meaning. One is lexical and comprises our semantic memory: our stored knowledge of the general meanings of words such as *queen, beauty*, or *come*. The other is grammatical and captures specific facts or events as located at particular moments in time, e.g. *He came here*, which involves references to a specific person and an event as located at a particular time and place. Sentential units of this latter type capture thoughts, which we can share and evaluate as true or false based on their referential content.

Both types of meaning exhibit principles of structural organization, yet these are crucially distinct. Lexical concepts are linked to each other through statistical associations (i.e., co-occurrences) as well as hierarchical relations (e.g. hyponymy and hypernymy, e.g., *queen → monarch*), while the links between words as occurring as parts of a sentence are grammatical (e.g., subject-predicate). Both lexical-associative and grammatical connectivity are essential for the organization of semantic structure in discourse. When producing one content word after another in word production tasks (e.g., verbal fluency tasks requiring the enumeration of animals), we are retrieving them from lexical semantic memory one by one as we go along, e.g., *mouse, cat, dog, camel*, reflecting a 'walk' in conceptual space. In natural speech, on the other hand, interleaved between these content words is the functional structure of grammar, which, together with the lexical concepts themselves, yields sentences and thoughts encoding references to scenes, objects, and events. Lexical selection among concepts, and references based on them, have to be integrated in order for any form of coherence to arise.

Deviance in lexical-semantic organization in psychosis has been previously studied psycholinguistically through the concept of anomalous 'spreading of semantic activations', often using semantic priming paradigms (Kuperberg et al., 2008; Pomarol-Clotet et al., 2008). Advances in natural language processing, on the other hand, have paved the way for the use of large computational Language Models (LM) to study semantic organization in terms of representations of words or sentences as vectors, named embeddings. Distances between these vectors can be calculated mathematically from their cosine angle. These distances are interpreted as semantic similarity, based on the distributional hypothesis (Baroni, 2013) that the co-occurrence patterns of a given word with other words captures its meaning. Static embeddings are those that represent each word according to the words it occurs with. Under such embeddings, words that occur with similar words will therefore get similar vectors, and they will lay in close areas in the vector space. Static embeddings have been adopted by the majority of previous computational semantic studies in psychosis (Corcoran et al., 2018; Corona-Hernández et al., 2022; Pauselli et al., 2018; Voppel et al., 2021).

However, reported diagnostic differences from such measures could arise from variations in superficial aspects of language organization such as syntactic frame selection (Yi et al., 2019) or sentence length (Hitczenko et al., 2021), rather than anomalies in semantic representations per se. There have also been inconsistencies across samples, languages, tasks, and processing pipelines, as revealed by several comparative studies (Iter et al., 2018; Just et al., 2019; Morgan et al., 2021; Parola et al., 2022). A specific case in point, addressed in the present study, is the inconsistency between repeated findings of a traditionally expected *de*crease in the mean semantic similarity between word pairs (Corcoran et al., 2018; Elvevåg et al., 2007; Iter et al., 2018; Morgan et al., 2021), and recent findings of a surprising *increase* in semantic similarity in two studies of first-episode psychosis (Alonso-Sánchez et al., 2022a; Pintos et al., 2022), and two studies of chronic psychosis samples (Parola et al., 2022; Voppel et al., 2021).

In normal connected speech, moreover, words never appear in isolation but form parts of grammatical structures with referential meaning, as illustrated above with *He came here*, which static semantic measures do not capture. Owing to their lack of sensitivity to either grammatical structure or extra-linguistic context, they do not capture particular occurrences of words within meaningful grammatical units, and they cannot target referential (as different from lexical-conceptual) meaning. This invites the utilization of contextual embeddings for studying semantic structure, due to their sensitivity to grammatical connectivity (Hill et al., 2014; Jawahar et al., 2019; Limisiewicz and Mareček, 2020). Both contextual and static (context-free) LMs successfully predict brain activity in people listening to a story during fMRI (Anderson et al., 2021; Goldstein et al., 2023; Kumar et al., 2023; Pasquiou et al., 2023), hence have some neurobiological validation. Yet, the two types of models also make partially different predictions and hence should be considered separately and comparatively in a study of meaning changes in psychosis. Previous findings using manual coding already support alterations not merely in conceptual but also referential meaning structure in psychosis samples across a range of typologically different languages (Çokal et al., 2022: Turkish; Çokal et al., 2018: English; Palominos et al., 2023: Chilean Spanish; Sevilla et al., 2018: peninsular Spanish). One notable repeated finding is an increase in pronouns compared to other types of noun phrases in psychosis groups (Çokal et al., 2022; Mackinley et al., 2021; Palominos et al., 2023; Tang et al., 2021). Crucially, pronouns are referential devices devoid of essentially any lexical-conceptual meaning, with regard to which static models of semantic similarity are validated. Syntactic complexity, too, has been reported as decreasing in psychosis (Barattieri di San Pietro et al., 2022; Ciampelli et al., 2023; Silva et al., 2022) and could bear on semantic structure as assessed with embeddings. This specifically applies to hierarchical syntactic complexity (the structure of phrases embedded in other phrases), which unlike non-hierarchical syntactical complexity (a linear sequence of words as measured, e.g., through sentence length) inherently captures complexity in meaning: the meaning of a complex phrase is computed *from* the meanings of the phrases embedded in it.

Meaning should not only be considered at both the lexical and grammatical levels, and contextually in the light of what is said before and after in a narrative, but the task of picture descriptions typically used for speech elicitation further invites considering the descriptions against the picture itself, which provides the frame of reference. This referential dimension of meaning can be targeted through a bimodal vision-language model, in which the picture described and the verbal description itself are separately vectorized, so that the semantic distance between the two can be scored (Radford et al., 2021). To address these desiderata, our aim here was to study contextual embeddings, at both word and sentence levels, in conjunction with static ones, and to contrast traditional cosine-similarity-based semantic metrics with the metric of 'perplexity', which quantifies the semantic unexpectedness of a word as occurring in its grammatical context, and a metric of bimodal semantic similarity between descriptions and pictures. To further justify the interrelatedness between syntactic complexity and meaning, we obtained metrics of hierarchical syntactic complexity and the ratio of pronouns. We assessed all metrics across three different picture descriptions and related them to clinical measures.

Our broad predictions were that due to grammatical organization being implicated in the use of contextual embeddings and differences in kind between meaning at the lexical and at the grammatical levels, semantic similarity metrics across the two types of models (static vs. contextual) would diverge. Perplexity in particular would increase in psychosis groups, indicating a loss in contextual semantic predictability, even if semantic similarity at the non-contextual level would increase, as reported in Alonso-Sánchez et al. (2022a). We also predicted an over-use of pronouns, and expected this overuse to impact on lexical-conceptual semantic similarity, as pronouns are devices lacking lexical-conceptual meaning. Finally, we expected hierarchical syntactic complexity to

correlate positively with contextual semantic models and perplexity, since contextual meaning depends on syntactic organization and more integrity in the latter (richer complexity) may enhance both semantic organization and predictability.

2. Methods

2.1. Data collection and clinical assessment

Ninety-four native English speakers from London, Ontario, Canada were recruited in this study and categorized as healthy controls (HC, $n = 29$), clinical high-risk (CHR, $n = 18$), first-episode psychosis (FEP, $n = 29$), and chronic schizophrenia (CS, $n = 18$), as a part of the Tracking Outcomes in Psychosis (TOPSY) study (https://clinicaltrials.gov/study/NCT02882204). Demographic data and clinical scores are shown in Table 1. The four groups were matched on gender and education, but not on age as expected, given the differences in illness stage (CHR = FEP = HC < CS). The FEP (Alonso-Sánchez et al., 2022b) and CHR samples (Jeon et al., 2021) are described in detail in our other reports focused on neuroimaging findings. In brief, FEP subjects had <2 weeks of lifetime antipsychotic exposure and in most cases were assessed in the first week of referral to the first-episode psychosis team. As such, the median dose of antipsychotic exposure, calculated by converting the various prescribed antipsychotic medication doses to a common equivalent on the basis of Defined Daily Dose (DDD) provided by the WHO Collaborating Centre for Drug Statistics and Methodology (https://www.whocc.no/atc_ddd_index_and_guidelines/guidelines/) and multiplying by the days of exposure to this dose, was <3 DDD-days in this sample. Only the data from FEP whose diagnosis remained stable (as schizophrenia, excluding those who had bipolar disorder or depressive psychosis) after 6 months of follow-up are included in this study. The CHR group included subjects with subthreshold psychosis (Attenuate Psychosis Syndrome or brief and limited intermittent psychosis (BLIPS)) as per the Brief Structured Interview for Psychosis-risk Syndromes (SIPS) with no prior exposure to antipsychotics ever in their lifetime. CS consisted of 18 subjects that were clinically stable on long-acting injectable medications with >3 years since illness onset and no recorded hospitalization in the past year and receiving community-based care from physicians affiliated to a first-episode clinic (PEPP, London Ontario). Importantly, CS subjects (and subjects in all other groups) were recruited regardless of the status of disorganization/thought disorder in their prior history, which was in order not to bias our sample towards language-related symptomatology. All diagnostic assessments were reviewed using a Best Estimate Procedure (Leckman et al., 1982) for clinical consensus (treating physician, a research psychiatrist and evaluators). All patients provided written informed consent as stipulated by the Research Ethics Committee of University of Western Ontario, London, Canada (ID 108268).

FEP and CS subjects were assessed with the Positive and Negative Syndrome Scale-8 items version (PANSS), with delusions (PANSS8P1), conceptual disorganization (PANSS8P2), hallucinatory behavior (PANSS8P3), blunted affect (PANSS8N1), passive/apathetic social withdrawal (PANSS8N4), lack of spontaneity/flow of conversation (PANSS8N6), mannerisms/posturing (PANSS8G5), and unusual thought content (PANSS8G9) (Opler et al., 2007). CHR subjects were assessed with the Scale of Prodromal Symptoms (SOPS) (Miller et al., 2003) for three positive symptoms: delusions (SOPS-S1, equivalent to PANSS8P1), conceptual disorganization (SOPS-S5, equivalent to PANSS8P2), hallucinatory behavior (SOPS-S4, equivalent to PANSS8P3). SOPS and PANSS are scored with the same scale (1–7) while HC subjects were all assessed as 1. All subjects were asked to describe three pictures from the Thematic Apperception Test (Murray, 1943) (described in supplementary materials) and were given one minute for each image. During the speech, if any participant would finish their descriptions in less than one minute, the interviewer would prompt them to speak more, and if they were continuing beyond one minute, the interviewer would interrupt them. This procedure makes the quantity of speech relatively similar across groups. No subject was excluded for not generating sufficient speech, and there were not significant group differences, as indicated by Kruskal–Wallis test with Dwass, Steel, Critchlow and Fligner all-pairs comparison post-hoc tests, on the number of words and utterances (as

Table 1
Demographic characteristics and clinical assessment of the subjects.

	Health control	Clinical high risk	First-episode psychosis	Chronic schizophrenia	Test	Statistics	p
Number	29	18	29	18	/	/	/
Age	22.00 (3.00)	21.00 (5.75)	22.00 (4.00)	27.00 (6.50)	Kruskal–Wallis	19.362	0.000***
Sex	24.14 %	22.22 %	27.59 %	27.78 %	Pearson's χ^2 test	0.245	0.970
Education	68.97 %	38.89 %	51.72 %	38.89 %	Pearson's χ^2 test	6.231	0.101
SES (parental)	3.0 (2.0)	4.0 (0.0)	4.0 (3.0)	4.0 (2.0)	Kendall's correlation	0.079	0.402
PANSS Missing	0	2	0	2	/	/	/
P1: Delusions	1.00 (0.00)	1.42 (2.25)	5.00 (2.00)	2.50 (3.00)	Kruskal–Wallis	58.095	0.000***
P2: Conceptual disorganization	1.00 (0.00)	0.00 (0.55)	3.00 (3.00)	1.00 (1.00)	Kruskal–Wallis	44.655	0.000***
P3: Hallucinatory behavior	1.00 (0.00)	3.00 (2.88)	5.00 (1.00)	3.00 (3.25)	Kruskal–Wallis	44.534	0.000***
N1: Blunted affect	1.00 (0.00)	/	2.00 (3.00)	1.00 (1.00)	Kruskal–Wallis	19.719	0.000***
N4: Passive/apathetic social withdrawal	1.00 (0.00)	/	3.00 (4.00)	1.00 (1.00)	Kruskal–Wallis	27.929	0.000***
N6: Lack of spontaneity and flow of conversation	1.00 (0.00)	/	1.00 (2.00)	1.00 (0.00)	Kruskal–Wallis	17.560	0.000***
G5: Mannerisms & posturing	1.00 (0.00)	/	1.00 (2.00)	1.00 (0.25)	Kruskal–Wallis	14.860	0.001***
G9: Unusual thought content	1.00 (0.00)	/	4.00 (2.00)	1.50 (1.25)	Kruskal–Wallis	44.792	0.000***
PANSSP	3.00 (0.00)	4.50 (3.21)	12.00 (4.00)	7.50 (5.25)	Kruskal–Wallis	61.398	0.000***
PANSSN	3.00 (0.00)	/	7.00 (6.00)	3.50 (1.75)	Kruskal–Wallis	32.979	0.000***
PANSSG	2.00 (0.00)	/	5.00 (4.00)	3.00 (2.00)	Kruskal–Wallis	44.972	0.000***
PANSSTOTAL	8.00 (0.00)	/	25.00 (9.00)	14.50 (9.50)	Kruskal–Wallis	58.585	0.000***
TLI Missing	2	0	0	1	/	/	/
Impoverishment of thought	0.00 (0.25)	0.50 (0.69)	0.25 (0.75)	0.25 (0.75)	Kruskal–Wallis	15.160	0.002**
Disorganization in thought	0.00 (0.25)	0.62 (0.69)	1.00 (1.25)	0.00 (0.50)	Kruskal–Wallis	18.437	0.000***

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Age was indicated by the median (Interquartile range, IQR). There was 1 missing value of age in CS, which were fulfilled with the mean age of the CS group. Missing values in other variables were excluded in the analysis. Sex was represented by the percentage of female subjects. Education was indicated by the percentage of subjects with over 12 years of education. Socioeconomic status (SES), PANSS scores, and the TLI scores were represented by the median (IQR). CHR group did not take PANSS test but SOPS where we obtained the scores for the three positive items but not for negative or general items so CHR was excluded for PANSS negative and PANSS general.

shown in supplementary materials). The recorded speech was transcribed by research assistants. Elicited speech was scored using the Thought Language Index (TLI), with scores for impoverishment of thought and for disorganization in thinking (Liddle et al., 2002). TLI global scores for three pictures were summed up per subject.

2.2. Semantic similarity analysis using four LMs

Our language-analytic pipeline is shown in Fig. 1. Every transcript was first segmented into meaningful units $(u_1, u_2, \cdots, u_n)$, which were encoded by LMs into embeddings $(e_1, e_2, \cdots, e_n)$, with n denoting the number of meaningful units. Meaningful units could either be words/tokens or utterances. Utterances were defined as syntactically independent units (neither necessarily nor sufficiently being *clauses*) providing new information to the discourse (Chapin et al., 2022; Çokal et al., 2022). Four of the authors worked together to manually divide the transcripts into utterances, with one single researcher with experiences in utterance division validating all divisions. Globally, in a segmented and embedded transcript $U = (e_1, e_2, \cdots, e_n)$, semantic similarity of U was defined as the averaged cosine similarity of every successive pairs, as shown in Eq. (1):

$$Similarity := \frac{1}{n-1} \sum_{i=1}^{n-1} cosine\_similarity(e_i, e_{i+1}) \quad (1)$$

First, transcripts were segmented into tokens using spaCy (3.4.2, en_core_web_sm, Montani et al., 2022). After removing all punctuations and stopwords (available in supplementary materials, nearly all pronouns were included as stopwords), the FastText model (Grave et al., 2018), pretrained on English data, was applied to encode all tokens. FastText is an LM returning a static embedding for each token regardless of the context. LMs like BERT (Devlin et al., 2019), by contrast, deliver contextual embeddings which are thought to capture aspects of hierarchical syntactic complexity, at certain layers of the transformer network (Jawahar et al., 2019). At the token level, the English uncased BERT was applied to tokenize the transcript and encode each token into an embedding. At utterance-level, two sentence embedding models were applied to encode each utterance with stopwords retained but punctuations removed. One of them was the FastText-based Infersent model (Conneau et al., 2017), and the other was SBERT pretrained on the pooled outputs from token-level contextual LMs to derive semantically meaningful sentence embeddings (Reimers and Gurevych, 2019). Infersent and SBERT were more reliable than averaged word-level embeddings for unsupervised sentence similarity evaluation (Sun et al., 2022). These two models brought our analyses from word/token level to utterance level for both context-free embeddings and contextual embeddings. Infersent encodes sentences with the bidirectional long short-term memory network (BiLSTM) structure, while SBERT replaced this with the Transformers-based encoder from BERT. BERT encoder captures the hierarchical syntactic structure of language (Jawahar et al., 2019), with a strong ability to learn long-term dependencies (Devlin et al., 2019), much better than classical LSTM and equivalent to an LSTM variant with inductive bias on syntax (Pei et al., 2020). SBERT is thus expected to encode more contextual information from the utterances than Infersent. However, despite operating at sentence-level with certain architectures used to embed contexts, both of these two were considered here to be unlike BERT in capturing little hierarchical grammar, thereby serving more to capture conceptual meaning as different from the contextual meaning of the utterance. This was mainly based on the following reasons: (1) The sentence embeddings are generated either by average- (SBERT) or max- (Infersent) pooling the token embeddings, which flattens the hierarchy of sentence structure. This point was independently supported in our results based on the lack of a predictive effect of syntactic depth on the semantic similarity scores derived from these two models (see results); (2) The embedding process for utterances is isolated, neglecting any contextual information from adjacent utterances; (3) During the supervised training of these models, the similarity labels are exclusively assigned based on the inherent semantic relationships between individual utterances, disregarding any contextual influences or dependencies.

Overall, this fourfold semantic similarity analysis covers both decontextualized lexical meaning and contextual referential meaning at the level of the meaningful utterance, where thoughts are expressed using grammar.

2.3. Semantic perplexity analysis

In addition to semantic similarity metrics, we assessed perplexity (PPL) of an utterance as a metric of the unexpectedness of its comprising units. PPL has been found to be a reliable speech coherence marker sensitive to cognitive decline, capturing meaning at the discourse level (Colla et al., 2022). Although PPL is not defined for masked LMs like BERT, a related metric, pseudo-perplexity (PPPL), has been proposed using a similar mathematical computation (Salazar et al., 2020). Formally, in a tokenized utterance $U := (t_1, t_2, \ldots, t_n)$, we defined the probability of token t_i as the log conditional probability of the utterance without this token $U_{\setminus i} := (t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n)$. The perplexity of the utterance was then defined as the exponential value of the negative mean value of the pseudo-loglikelihood scores provided by summing the conditional log probabilities of all tokens in the utterance, as shown by Eq. (2):

$$perplexity(U_t) := Exp\left(-\frac{1}{n} \sum_{i=0}^{n} \log P_{LM}(t_i | U_{\setminus i}) \right) \quad (2)$$

We computed the PPPL scores for every utterance in the transcript using BERT, and averaged them to obtain a single PPPL score for every subject.

2.4. Multimodal similarity analysis: from image to text

As a final LM, we used CLIP (Radford et al., 2021), a vision-language pretrained model designed to encode both an image and a text and to quantify the similarity between them. For each subject, we computed the cosine similarity between an image and every utterance describing it using CLIP with visual transformer (ViT-B/32) as its backbone, and averaged the image-utterance similarity scores.

2.5. Syntactic measures with impact on semantic structure

To shed light on the relationship between elements of referential structure and conceptually based semantic evaluations, we computed the ratio of pronouns and hierarchical syntactic complexity. The required part-of-speech tagging was carried out with the identical model from spaCy. For hierarchical syntactic complexity, a constituency parser (benepar_en3) (Kitaev and Klein, 2018) was utilized with spaCy. Punctuations were kept for parsing the trees but removed from the trees for sequential analysis. The constituency parse was represented as a directed acyclic graph with the source (sentence) node at top representing the whole utterance, target nodes at bottom representing the part-of-speech tags of every token, and stem nodes in between representing phrases. Though some previous studies of syntax in psychosis such as Ciampelli et al. (2023) have constructed syntax trees in Greibach Normal Form, allowing a parent node to have more than two children nodes, we converted the Greibach Normal Form to Chomsky Normal Form, which allows a parent node to have at most two children nodes, using the *nltk* package. For instance, in Fig. 2(A) and (B) shows the syntactic structure of a noun phrase, *the red apple*, in the Greibach and Chomsky normal forms, respectively. As illustrated here, the omission of intermediate phrasal structures, such as the internal NP structure in this case, abstracts from some layers of the hierarchical syntactic complexity. In our analysis, forms of tokens were not included in syntactic trees, as

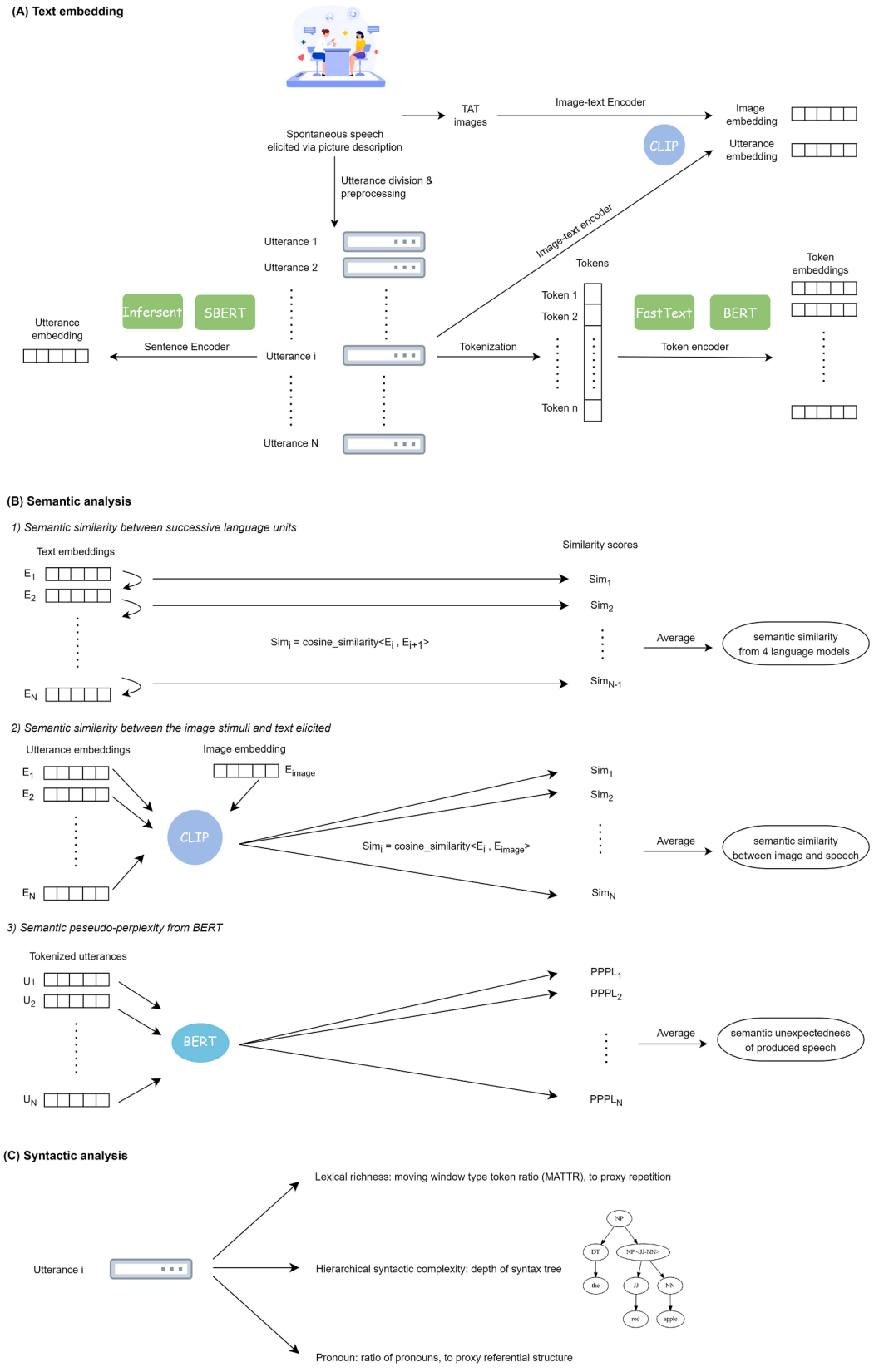

Fig. 1. Workflow for language analysis pipelines, with (A) transcript preprocess and encoding, (B) semantic analysis for successive textual similarity, image-text similarity, and textual perplexity, and (C) syntactic analysis.

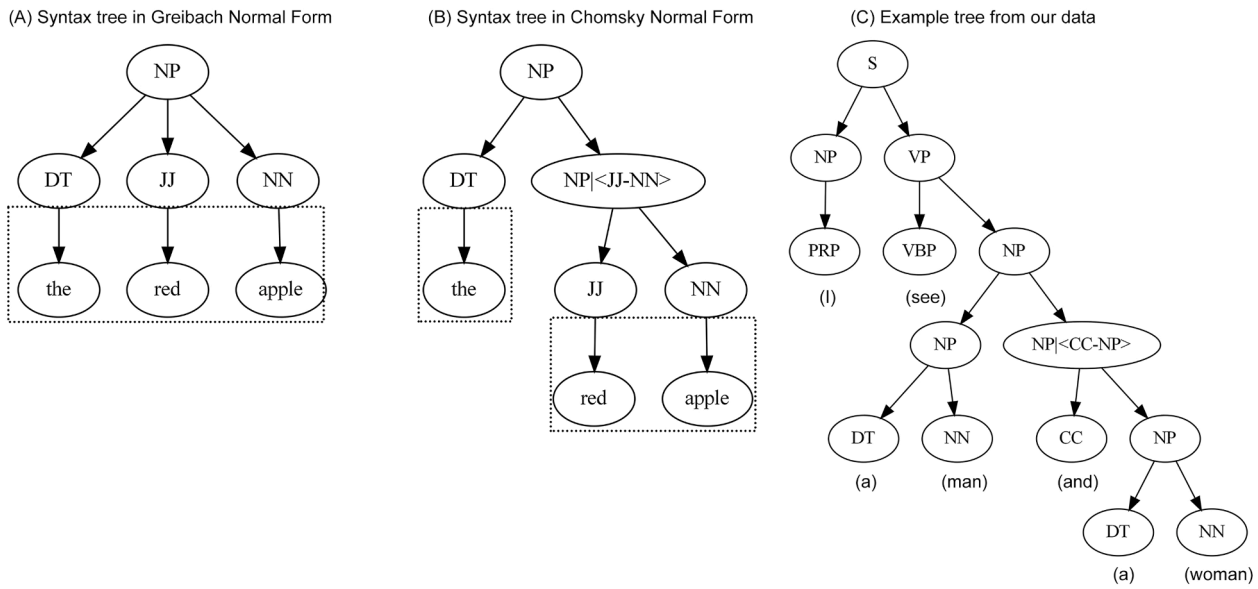

Fig. 2. Examples of constituency parsing. Tokens in dotted boxes are only for demonstration and excluded in actual analyses. (A) Syntactic structure of *The red apple* in Greibach Normal Form; (B) Syntactic structure of *The red apple* in Chomsky Normal Form; (C) Syntactic structure of a real example from our data *I see a man and woman*. The meanings of node labels appearing in the trees: S: sentence, NP: noun phrase, VP: verb phrase, DT: determiner, JJ: adjective, NN: noun (singular or mass), PRP: personal pronoun, VBP: verb (non-3rd person singular present), CC: coordinating conjunction. $Node_i \mid \langle Node_{i-0},\ Node_{i-1} \rangle$ refers to the phrase $Node_i$ being comprised of its two child nodes.

shown in Fig. 2(C). Syntactic depth was defined as the maximum number of edges that need to be traversed to reach from the source node (sentence node) to the furthest target node (token node). For example, in Fig. 2(C), the depth of syntax tree is five, from S to NN (woman) or DT (a).

Repetitions and other dysfluencies have been recognized in schizophrenia (Çokal et al., 2019) and were also observed in our study (see results). Repeated expressions could exert a substantial influence on semantic similarity (in the direction of high semantic similarities). To control for this, we used the moving-averaged type-token-ratio (MATTR) with a window size of 25 (Covington and McFall, 2010) as a proxy for repetition, as more repetitions are expected to lead to less unique words and thus lower MATTR.

2.6. Group comparisons

Each participant described three pictures, hence there were three correlated observations of language measures per participant. Generalized estimating equation (GEE) models were applied to estimate population-averaged effects, with Gaussian response distribution, exchangeable correlation structure, and log link function. Language measures served as response variables while age and group as independent variables (with HC as the reference category). Furthermore, for comparisons within the clinical groups, we used the same method but set FEP as the reference category, with all HC data excluded. MATTR was defined as an offset variable if there was a significant relationship between MATTR and the response linguistic variable (as described in 2.7). If there was no significant relationship, MATTR was not included in the GEE model. False discovery rate (FDR) was applied to correct p values for each pathological group within two domains, semantic and syntactic, and is reported as q values in this paper. To better fit the GEE model, we first log-transformed the BERT PPPL scores and syntactic depth to balance distribution. BERT PPPL scores were further Box–Cox transformed to fit the Gaussian distribution (Box and Cox, 1964). Deviance goodness-of-fit test was applied to assess how well the GEE models fit the observed data.

2.7. Relationship among semantic measures, syntactic measures and clinical scores

Identical GEE models were applied to predict semantic features from syntactic features, to explore their association. All subjects were included in the analysis with group treated as covariate in the GEE models to regress out the effect of group variance. Furthermore, to investigate which clinical aspect of cognition these language measures relate to, generalized linear models (GLM) were applied to predict, from task-averaged linguistic measures: (1) the score of each PANSS items. For CHR subjects, the positive items were represented by corresponding SOPS scores; (2) PANSS positive (PANSSP, P1+P2+P3), negative (PANSSN, N1+N4+N6), general (PANSSG, G5+G9), and total scores; and (3) two scores from Thought and Language Index (TLI), the global impoverishment of thought score, and the global disorganization in thought score. PANSS scores fit the Tweedie distribution while TLI scores fit the Gaussian distribution. Domain-wise correction for p values with FDR for each pathological group was applied. The PPPL scores and syntactic depth were not transformed here and deviance goodness-of-fit test was applied to assess how well the GLM models fit the observed data.

3. Results

3.1. Group differences

All GEE and GLM models in this study fit the data well (deviance goodness-of-fit: all $p > 0.05$). As shown in Table 2, compared to HC, static semantic similarity was higher in CHR (FastText) as well as FEP groups (FastText, Infersent, SBERT). CLIP-based multimodal (picture-description) alignment was lower in both of these groups. Only FEP had higher contextual unexpectedness (PPPL scores). The CS group had no notable aberrations that survived our a priori statistical threshold. Significant changes in syntactic measures were observed only in FEP, with lower MATTR indicating more repetition, a higher ratio of pronouns, and greater syntactic depth. MATTR was included in the GEE models for the sematic similarity scores from Infersent and SBERT, the PPPL scores, and syntactic depth, due to significant predictive effects observed (see

Table 2
Comparisons of the three pathological groups to health controls on language measures.

Group	Domain	Feature	*B*	*se*	*z*	*p*	*q*
CHR	Semantics	FastText	0.085	0.030	2.876	0.004**	0.012*
		Infersent	0.074	0.040	1.873	0.061	0.122
		SBERT	0.031	0.046	0.659	0.510	0.510
		BERT	0.013	0.014	0.903	0.367	0.510
		CLIP	−0.020	0.006	−3.150	0.002**	0.010*
		BERT_pppl	0.022	0.030	0.728	0.467	0.510
	Syntax	MATTR	−0.016	0.013	−1.210	0.226	0.526
		PRON	0.048	0.052	0.934	0.350	0.526
		Depth	0.004	0.021	0.178	0.859	0.859
FEP	Semantics	FastText	0.120	0.032	3.741	0.000***	0.001***
		Infersent	0.153	0.034	4.538	0.000***	0.000***
		SBERT	0.092	0.040	2.281	0.023*	0.034*
		BERT	−0.024	0.016	−1.450	0.147	0.147
		CLIP	−0.016	0.008	−2.157	0.031*	0.037*
		BERT_pppl	0.059	0.024	2.481	0.013*	0.026*
	Syntax	MATTR	−0.030	0.015	−2.007	0.045*	0.045*
		PRON	0.155	0.043	3.626	0.000***	0.001***
		Depth	0.036	0.018	2.024	0.043*	0.045*
CS	Semantics	FastText	0.042	0.025	1.714	0.087	0.260
		Infersent	0.023	0.037	0.631	0.528	0.622
		SBERT	0.024	0.049	0.493	0.622	0.622
		BERT	−0.023	0.019	−1.180	0.238	0.476
		CLIP	−0.007	0.007	−0.896	0.370	0.555
		BERT_pppl	0.058	0.029	2.016	0.044*	0.260
	Syntax	MATTR	−0.026	0.013	−2.062	0.039*	0.118
		PRON	−0.008	0.053	−0.145	0.885	0.885
		Depth	0.021	0.020	1.042	0.297	0.446

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. HC is set as the reference category. We only reported the coefficient (*B*), standard error (*se*), z score (z), *p* values before correction (*p*), and *p* values after correction (*q*) for every group here to make the table concise, with the complete table of results available in supplementary materials. Semantics: semantic similarity based on five language models and the perplexity scores based on BERT. Syntax: moving-window averaged type token ratio (MATTR), ratio of pronouns (PRON), syntactic depth (Depth). Same below.

Methods and Fig. 3). As shown in Table 3, with all HC data excluded, compared to FEP, the CHR group showed no changes that survived thresholds, while the CS group had lower semantic similarity based on Infersent and a lower ratio of pronouns.

3.2. Predicting semantic variables using syntactic variables

In line with predictions, a relation transpired between syntactic measures (the pronoun ratio and syntactic complexity), on the one hand, and semantic similarity and perplexity measures, on the other (see Fig. 3). In particular, higher syntactic depth predicted lower BERT semantic similarity scores and less perplexity. More pronouns predicted higher semantic similarity scores from all unimodal LMs except BERT, despite a trend-level relationship between more pronouns and lower semantic similarity with BERT ($q = 0.073$). They also predicted increased perplexity and decreased semantic similarity with the bimodal CLIP model.

3.3. Relations to clinical scores

As shown in Fig. 4, the general syndromic changes as measured by the total PANSS scores was only predicted by perplexity, with higher perplexity predicting higher burden of unusual thought content and delusions. In addition, lower CLIP scores predicted higher conceptual disorganization while higher context-free word-based semantic similarity (from FastText) predicted increasing lack of spontaneity and flow of conversation. Lower syntactic depth predicted worse passive/apathetic social withdrawal. As for the TLI scores, increasing

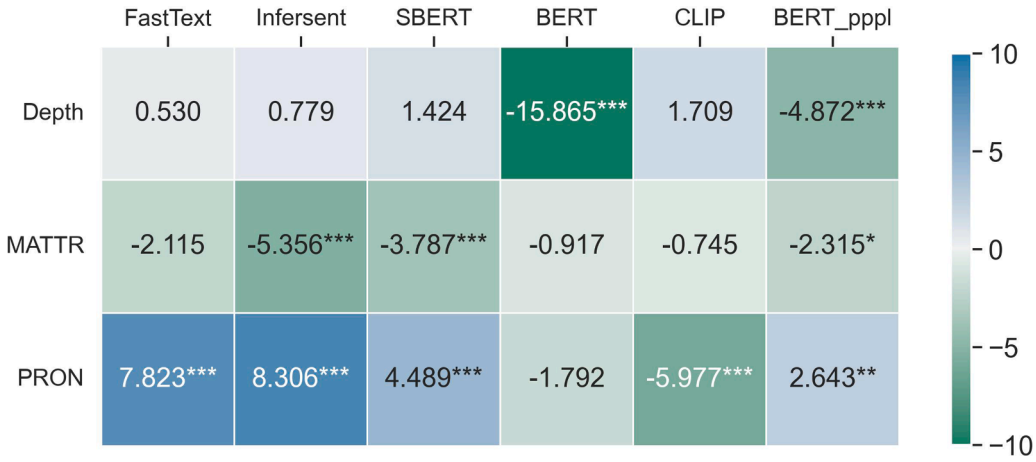

Fig. 3. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Prediction of semantic measures using syntactic measures. Blue squares indicated positive effects while the green squares indicated negative effects. Displayed numbers represent *z* scores. The larger the *z* score, the denser the color.

Table 3
Comparisons of clinical high risk and chronic schizophrenia to first-episode psychosis on language measures.

Group	Domain	Feature	*B*	*se*	*z*	*p*	*Q*
Clinical High Risk	Semantics	FastText	−0.041	0.045	−0.913	0.361	0.434
		Infersent	−0.080	0.047	−1.716	0.086	0.258
		SBERT	−0.063	0.054	−1.163	0.245	0.368
		BERT	0.036	0.016	2.257	0.024*	0.144
		CLIP	−0.004	0.008	−0.450	0.653	0.653
		BERT_pppl	−0.036	0.028	−1.293	0.196	0.368
	Syntax	MATTR	0.015	0.018	0.831	0.406	0.406
		PRON	−0.106	0.054	−1.972	0.049*	0.146
		Depth	−0.030	0.022	−1.401	0.161	0.242
Chronic Schizophrenia	Semantics	FastText	−0.069	0.038	−1.813	0.070	0.209
		Infersent	−0.131	0.045	−2.929	0.003**	0.020*
		SBERT	−0.082	0.057	−1.438	0.150	0.301
		BERT	−0.003	0.021	−0.152	0.879	0.908
		CLIP	0.006	0.009	0.634	0.526	0.790
		BERT_pppl	−0.003	0.026	−0.115	0.908	0.908
	Syntax	MATTR	0.005	0.017	0.304	0.761	0.830
		PRON	−0.141	0.054	−2.612	0.009**	0.027*
		Depth	−0.005	0.022	−0.215	0.830	0.830

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. First-episode psychosis is set as the reference category. We only reported the coefficient (*B*), standard error (*se*), z score (z), *p* values before correction (*p*), and *p* values after correction (*q*) for every group here to make the table concise, with the complete table of results available in supplementary materials.

impoverishment of thought was predicted by higher semantic similarity measured with FastText and Infersent and higher MATTR. More disorganization in thought was predicted by lower CLIP scores and a higher ratio of pronouns.

4. Discussion

The aim of this study was to use both context-free and contextual LMs to expound the organization of meaning in psychosis at two different levels of linguistic organization, lexical and grammatical, unimodally and bimodally, and the possible impact on semantic measures of three factors: the ratio of pronouns, lexical diversity (MATTR), and hierarchical syntactic complexity (syntactic depth). The main findings can be summarized as follows: (a) In early psychosis (CHR and FEP), there is an increase in static semantic similarity that relates to impoverished thinking; (b) in acute stages, i.e. FEP, this is accompanied by insignificantly reduced context-dependent similarity (both the context from prior utterances and the ground truth from a different modality) and significantly more semantic uncertainty as measured by perplexity, which related to disorganized thinking, unusual thought content, and delusions; (c) in the stable state of chronic established schizophrenia (CS), neither of these deficits are apparent; (d) semantic metrics associate with syntactic factors.

Use of computational language models in psychosis dates back to pioneering work by Elvevåg et al. (2007), who used Latent Semantic Analysis (LSA) to quantify the 'coherence' of a text via semantic similarity. Ever since, numerous studies have pursued a similar conceptualization of semantic similarity as measuring 'coherence', and of decreased semantic similarity as relating to formal thought disorder (FTD). In line with this, decreases in mean semantic similarity in both psychosis and high-risk samples have been expected and widely reported, in studies that used static embeddings such as those produced by LSA (Bedi et al., 2015; Corcoran et al., 2018) or Word2Vec (Figueroa-Barra et al., 2022; Iter et al., 2018; Morgan et al., 2021), or else contextual ones such as those derived from BERT (Tang et al., 2021). Conversely, there is the divergent finding of an unsuspected increase in semantic similarity using context-free embeddings from GloVe (Alonso-Sánchez et al., 2022a), which we replicated here using FastText. Given the conceptualization of these metrics as measuring coherence and FTD, higher semantic similarity is a surprise. Our starting point here, however, was different and focused on the organization of the semantic space, which is what the models represent, and its relation to linguistic structure. Under the de-grammaticalized vision of meaning captured by static embeddings, higher semantic similarity indicates that, when producing word after word, people with FEP are navigating within a more constrained semantic space, where distances travelled between lexical concepts are reduced. The same phenomenon is seen in the same group on the sentential level using sentential embeddings from Infersent and SBERT, which encode more formal syntactic information than bag-of-word models (e.g., FastText), but less than BERT (Chrupała and Alishahi, 2019), likely because the hierarchical syntactic structures are flattened in the pooling operation used in these models. Static models, however, are only selective windows, leaving much of the complexity aside, and when moving to the grammatical level using contextual embeddings, a very different pattern emerges: lower semantic similarity with the contextual model BERT, correlating with hierarchical syntactic complexity, and with the bimodal model CLIP, which evaluates speech against the referential context; accompanied by increasing perplexity with regards to how words appear in grammatically structured utterances.

The view of semantic similarity as measuring coherence therefore falls short: rather there is a dual pattern of increased context-free but reduced contextual semantic similarity. Put differently, there is a tightening of the lexical-conceptual semantic space (as indexed by reduced distances between lexical representations), but apparent loosening when words are considered in their grammatical context using BERT. The negative association, in both the cases of semantic similarity and perplexity, between this loosening and syntactic complexity is telling, pointing to a specific role of grammar in the organization of meaning and suggesting a lesser grip of grammar on meaning at the referential level relevant to cohesive discourse. It appears as if perplexity and semantic expansion take place when hierarchical syntax cannot retain its complexity, suggesting that grammar, when intact, could act as a mechanism of semantic control.

The increase in pronouns seen in FEP echoes similar findings in other samples and languages, and contributes to profiling a shift in the referential usage of language through grammar. Pronouns encode no descriptive lexical-conceptual content whatsoever, i.e., they do not represent their referents through a lexical concept that needs to be remembered and retrieved. They are in this sense purely grammatical devices with only a referential (but no representational) function. Why, then, would the increase of such 'shortcuts' to reference predict changes in the organization of the semantic space? We suggest that the answer lies in the fact that pronouns are deictic devices, and as such can only be used to pick out referents provided in the immediate speech contexts. They therefore do not allow to cross large semantic distances, confining

Fig. 4. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Prediction of PANSS and TLI scores using language measures. Blue squares indicated positive effects while the green squares indicated negative effects. Displayed numbers represent z scores. The larger the z score, the denser the color.

us to a space already spanned by the concepts of the previous discourse or else the immediate non-linguistic context (such as the picture). As purely referential devices failing to *represent* the object of reference through some descriptive feature, pronouns thus bypass the need for further conceptual-semantic processing. Based on this it is natural to expect that an increase in the pronoun ratio is related, or an overt expression of, a mechanism affecting the density of the `packing' of the semantic space.

The new concept of a shrinking conceptual-semantic space should be considered in a broader (neuro-)biological context. In most biological systems, the notion of clustering (connectedness to the neighborhood) is balanced with the concept of integration (reaching out to elements at greater distances), with a 'small world' topological organization being the consequence of balancing these two (Lord et al., 2017). From this perspective, the apparent loss of semantic relatedness underlying the construct of 'loosening associations' could arise from an imbalance between integration (context-driven link between seemingly unrelated words) and segregation (clustering among related words) of concepts in schizophrenia. In our model and data, it is the lexical-conceptual space that tightens (comparable to higher segregation), while referential meaning as carried by the grammatical level loosens (comparable to reduced integration), which is in line with numerous studies of 'referential anomalies' in psychosis (Çokal et al., 2018; Rochester and Martin, 1979; Sevilla et al., 2018).

Changes in semantic similarity have traditionally not only been conceptualized as measures of 'coherence' but also as measures of FTD. Associations with clinical measures of FTD have been inconsistent (Parola et al., 2022), however, and some studies have reported a lack of correlation with FTD scores based on the decreases (Tang et al., 2021) or increases (Alonso-Sánchez et al., 2022a) in semantic similarity. In the present study, impoverishment of thought was predicted by an increase in similarity scores from static embeddings, as well as lower lexical richness (MATTR), both of which resonate with the idea of a deflating lexical-conceptual semantic space. In turn, disorganization in thought, reflecting looseness and peculiar language usage, was predicted by two language measures on the side of referential-grammatical meaning, namely lower CLIP-based similarity and overuse of pronouns. These correlations are independent evidence for the need to distinguish between conceptual and referential semantics, and the clinical significance of this distinction. Disorganization and impoverishment are thought to be orthogonal phenomena (Liddle et al., 2002). They also further support our previous argument that deflation in conceptual semantic space as evidenced by measures like FastText similarity can live in harmony with the expansion in referential semantic space as evidenced by measures like CLIP similarity, with both these aspects jointly constituting 'loosening of associations'. In line with this idea, PPPL predicted the overall worsening of the symptoms, as well as delusions and unusual thought content, while CLIP-based similarity predicted conceptual disorganization. These two language measures, due to their relationship with the positive symptoms, could be considered as potential markers for state changes during the disease course.

4.1. Strengths and limitations

To the best of our knowledge, this is the first study that comparatively examines semantic changes in psychosis by utilizing different LMs to differentiate conceptual meanings and grammar-mediated referential meanings. Additionally, it is the first study to explore the significance of CLIP scores and perplexity metrics in assessing semantic changes in psychosis. Another strength lies in the fact that the FEP group involved was drug-naïve, while the CS group was stable with low PANSS and TLI scores and years of medication. It is also noteworthy that our study had a relatively small sample size, particularly for CHR and CS, with a number of different analyses. In addition, we did not assess IQ formally but one of our exclusion criteria was having a diagnosis of Intellectual Disability. This effectively restricted the range of variations in verbal IQ. Lower IQ is seen as a feature of schizophrenia (Kremen et al., 2001), matching/controlling for which will reduce the variance related to the illness per se in case-control studies. Future work needs to explore whether semantic structural changes would differ under alternative conditions, such as free conversational speech. Subsequent investigations should aim to validate our findings on a larger dataset, employing speech elicited under various conditions.

Data and code availability

Transcripts used for this study as well as anonymised clinical scores, are available from LP (write to lpalaniy@uwo.ca) upon reasonable request within the stipulations laid by The Research Ethics Committee of University of Western Ontario, London, Canada (Project ID: 108268; Most recent review reference: 2022-108268-71496; Study Title: The Pathophysiology of Thought Disorder in Psychosis (TOPSY)). Feature extraction, statistical analyses, and visualization were carried out using Python 3.9.12 and relevant packages. Scripts are available from the corresponding author upon reasonable request. Feature extraction, statistical analyses, and visualization were carried out using Python 3.9.12 and relevant packages. All feature extraction was conducted in April 2023. Scripts are available from the corresponding author upon reasonable request.

Funding

His research was supported by the Ministerio de Ciencia, Innovación y Universidades (MCIU) and the Agencia Estatal de Investigación (AEI) (grant PID2019-105241GB-I00/AEI/10.13039/501100011033 to WH), the grant TRUSTING, HORIZON-HLTH-2022-STAYHLTH-01, grant nr. 101080251-2 (to WH), China Scholarship Council (grant 202108390062 to RH), the Department of Science and Technology of Guangdong Province (grant 112175605105 to WH and RH), and the National Agency for Research and Development (ANID), Scholarship Program, Becas Chile 2019, Postdoctoral Fellow 74200048 (MA) (to MFAS). The data acquisition for this study was funded by CIHR Foundation Grant (FDN 154296) to LP and was supported by the Canada First Excellence Research Fund to BrainSCAN, Western University (Imaging Core); Innovation fund for Academic Medical Organization of Southwest Ontario; Bucke Family Fund, The Chrysalis Foundation and The Arcangelo Rea Family Foundation (London, Ontario). Compute Canada Resources (Application no. 1530) were used in the storage and analysis of imaging data. LP acknowledges research support from the Canada First Research Excellence Fund, awarded to the Healthy Brains, Healthy Lives initiative at McGill University (New Investigator Supplement); Monique H. Bourgeois Chair in Developmental Disorders and Graham Boeckh Foundation (Douglas Research Centre, McGill University) and a salary award from the Fonds de recherche du Quebec-Santé (FRQS).

CRediT authorship contribution statement

Rui He: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Conceptualization. **Claudio Palominos:** Writing – review & editing, Methodology, Conceptualization. **Han Zhang:** Methodology, Conceptualization. **Maria Francisca Alonso-Sánchez:** Writing – review & editing, Resources, Data curation, Conceptualization. **Lena Palaniyappan:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **Wolfram Hinzen:** Writing – review & editing, Writing – original draft, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

LP reports personal fees for serving as chief editor from the Canadian Medical Association Journals, speaker/consultant fee from Janssen Canada and Otsuka Canada, SPMM Course Limited, UK, Canadian Psychiatric Association; book royalties from Oxford University Press; investigator-initiated educational grants from Janssen Canada, Sunovion and Otsuka Canada outside the submitted work. All other authors report no relevant conflicts.

Acknowledgments

We appreciate all the participants and their families for the time and effort to contribute to this study. We also acknowledge Michael Mackinley, Jenny Chan, and Sabrina Ford of Western University for their support in preparing the dataset.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.psychres.2024.115752.

References

Alonso-Sánchez, M.F., Ford, S.D., MacKinley, M., Silva, A., Limongi, R., Palaniyappan, L., 2022a. Progressive changes in descriptive discourse in First Episode Schizophrenia: a longitudinal computational semantics study. Schizophr 8, 1–9. https://doi.org/10.1038/s41537-022-00246-8.

Alonso-Sánchez, M.F., Limongi, R., Gati, J., Palaniyappan, L., 2022b. Language network self-inhibition and semantic similarity in first-episode schizophrenia: a computational-linguistic and effective connectivity approach. Schizophr. Res. https://doi.org/10.1016/j.schres.2022.04.007.

Anderson, A.J., Kiela, D., Binder, J.R., Fernandino, L., Humphries, C.J., Conant, L.L., Raizada, R.D.S., Grimm, S., Lalor, E.C., 2021. Deep artificial neural networks reveal a distributed cortical network encoding propositional sentence-level meaning. J. Neurosci. 41, 4100–4119. https://doi.org/10.1523/JNEUROSCI.1152-20.2021.

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

Barattieri di San Pietro, C., Barbieri, E., Marelli, M., de Girolamo, G., Luzzatti, C., 2022. Processing argument structure and syntactic complexity in people with schizophrenia spectrum disorders. J. Commun. Disord. 96, 106182 https://doi.org/10.1016/j.jcomdis.2022.106182.

Baroni, M., 2013. Composition in distributional semantics. Lang. Linguist. Compass 7, 511–522. https://doi.org/10.1111/lnc3.12050.

Bedi, G., Carrillo, F., Cecchi, G.A., Slezak, D.F., Sigman, M., Mota, N.B., Ribeiro, S., Javitt, D.C., Copelli, M., Corcoran, C.M., 2015. Automated analysis of free speech predicts psychosis onset in high-risk youths. npj Schizophr. 1, 1–7. https://doi.org/10.1038/npjschz.2015.30.

Bleuler, E., 1911. Dementia Praecox, Oder Gruppe der Schizophrenien. Deuticke.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J. R. Stat. Soc. Ser. B (Methodological) 26, 211–252.

Chapin, K., Clarke, N., Garrard, P., Hinzen, W., 2022. A finer-grained linguistic profile of Alzheimer's disease and mild cognitive impairment. J. Neurolinguist. 63, 101069 https://doi.org/10.1016/j.jneuroling.2022.101069.

Chrupała, G., Alishahi, A., 2019. In: Correlating Neural and Symbolic Representations of Language. Presented at the Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 2952–2962. https://doi.org/10.18653/v1/P19-1283.

Ciampelli, S., de Boer, J.N., Voppel, A.E., Corona Hernandez, H., Brederoo, S.G., van Dellen, E., Mota, N.B., Sommer, I.E.C., 2023. Syntactic network analysis in schizophrenia-spectrum disorders. Schizophr. Bull. 49, S172–S182. https://doi.org/10.1093/schbul/sbac194.

Çokal, D., Palominos-Flores, C., Yalınçetin, B., Türe-Abacı, Ö., Bora, E., Hinzen, W., 2022. Referential noun phrases distribute differently in Turkish speakers with schizophrenia. Schizophr. Res. https://doi.org/10.1016/j.schres.2022.06.024. S0920-9964(22)00259-6.

Çokal, D., Sevilla, G., Jones, W.S., Zimmerer, V., Deamer, F., Douglas, M., Spencer, H., Turkington, D., Ferrier, N., Varley, R., Watson, S., Hinzen, W., 2018. The language profile of formal thought disorder. npj Schizophr. 4, 1–8. https://doi.org/10.1038/s41537-018-0061-9.

Çokal, D., Zimmerer, V., Turkington, D., Ferrier, N., Varley, R., Watson, S., Hinzen, W., 2019. Disturbing the rhythm of thought: speech pausing patterns in schizophrenia, with and without formal thought disorder. PLoS One 14. https://doi.org/10.1371/journal.pone.0217404.

Colla, D., Delsanto, M., Agosto, M., Vitiello, B., Radicioni, D.P., 2022. Semantic coherence markers: the contribution of perplexity metrics. Artif. Intell. Med. 134, 102393 https://doi.org/10.1016/j.artmed.2022.102393.

Conneau, A., Kiela, D., Schwenk, H., Barrault, L., Bordes, A., 2017. Supervised learning of universal sentence representations from natural language inference data. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Presented at the EMNLP 2017. Association for Computational Linguistics, Copenhagen, Denmark, pp. 670–680. https://doi.org/10.18653/v1/D17-1070.

Corcoran, C.M., Carrillo, F., Fernández-Slezak, D., Bedi, G., Klim, C., Javitt, D.C., Bearden, C.E., Cecchi, G.A., 2018. Prediction of psychosis across protocols and risk cohorts using automated language analysis. World Psychiatry 17, 67–75. https://doi.org/10.1002/wps.20491.

Corona-Hernández, H., de Boer, J.N., Brederoo, S.G., Voppel, A.E., Sommer, I.E.C., 2022. Assessing coherence through linguistic connectives: analysis of speech in patients with schizophrenia-spectrum disorders. Schizophr. Res. https://doi.org/10.1016/j.schres.2022.06.013. S0920-9964(22)00248–1.

Covington, M.A., McFall, J.D., 2010. Cutting the Gordian knot: the moving-average type–token ratio (MATTR). J. Quant. Linguist. 17, 94–100. https://doi.org/10.1080/09296171003643098.

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2019. BERT: pre-training of deep bidirectional transformers for language understanding. https://doi.org/10.48550/arXiv.1810.04805.

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

Figueroa-Barra, A., Del Aguila, D., Cerda, M., Gaspar, P.A., Terissi, L.D., Durán, M., Valderrama, C., 2022. Automatic language analysis identifies and predicts schizophrenia in first-episode of psychosis. Schizophrenia 8, 1–8. https://doi.org/10.1038/s41537-022-00259-3.

Goldstein, A., Ham, E., Nastase, S.A., Zada, Z., Grinstein-Dabus, A., Aubrey, B., Schain, M., Gazula, H., Feder, A., Doyle, W., Devore, S., Dugan, P., Friedman, D., Brenner, M., Hassidim, A., Devinsky, O., Flinker, A., Levy, O., Hasson, U., 2023. Correspondence between the layered structure of deep language models and temporal structure of natural language processing in the human brain. https://doi.org/10.1101/2022.07.11.499562.

Grave, E., Bojanowski, P., Gupta, P., Joulin, A., Mikolov, T., 2018. Learning word vectors for 157 languages. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). Presented at the Language Resources and Evaluation Conference.

Hill, F., Cho, K., Jean, S., Devin, C., Bengio, Y., 2014. Not all neural embeddings are born equal. https://doi.org/10.48550/arXiv.1410.0718.

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

Iter, D., Yoon, J., Jurafsky, D., 2018. Automatic detection of incoherent speech for diagnosing schizophrenia. In: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic. Presented at the Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic. Association for Computational Linguistics, New Orleans, LA, pp. 136–146. https://doi.org/10.18653/v1/W18-0615.

Jawahar, G., Sagot, B., Seddah, D., 2019. What does BERT learn about the structure of language?. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Presented at the ACL 2019. Association for Computational Linguistics, Florence, Italy, pp. 3651–3657. https://doi.org/10.18653/v1/P19-1356.

Jeon, P., Limongi, R., Ford, S.D., Branco, C., Mackinley, M., Gupta, M., Powe, L., Théberge, J., Palaniyappan, L., 2021. Glutathione as a molecular marker of functional impairment in patients with at-risk mental state: 7-Tesla 1H-MRS study. Brain Sci. 11, 941. https://doi.org/10.3390/brainsci11070941.

Just, S., Haegert, E., Kořánová, N., Bröcker, A.-L., Nenchev, I., Funcke, J., Montag, C., Stede, M., 2019. Coherence models in schizophrenia. In: Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology. Presented at the CLPsych 2019. Association for Computational Linguistics, Minneapolis, Minnesota, pp. 126–136. https://doi.org/10.18653/v1/W19-3015.

Kitaev, N., Klein, D., 2018. Constituency Parsing with a Self-Attentive Encoder. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Presented at the ACL 2018. Association for Computational Linguistics, Melbourne, Australia, pp. 2676–2686. https://doi.org/10.18653/v1/P18-1249.

Kremen, W.S., Seidman, L.J., Faraone, S.V., Tsuang, M.T., 2001. Intelligence quotient and neuropsychological profiles in patients with schizophrenia and in normal volunteers. Biol. Psychiatry 50 (6), 453–462. https://doi.org/10.1016/s0006-3223(01)01099-x.

Kumar, S., Sumers, T.R., Yamakoshi, T., Goldstein, A., Hasson, U., Norman, K.A., Griffiths, T.L., Hawkins, R.D., Nastase, S.A., 2023. Shared functional specialization in transformer-based language models and the human brain. bioRxiv 2022.06.08.495348. https://doi.org/10.1101/2022.06.08.495348.

Kuperberg, G.R., Lakshmanan, B.M., Greve, D.N., West, W.C., 2008. Task and semantic relationship influence both the polarity and localization of hemodynamic modulation during lexico-semantic processing. Hum. Brain Mapp. 29, 544–561. https://doi.org/10.1002/hbm.20419.

Leckman, J.F., Sholomskas, D., Thompson, W.D., Belanger, A., Weissman, M.M., 1982. Best estimate of lifetime psychiatric diagnosis: a methodological study. Arch. Gen. Psychiatry 39, 879–883. https://doi.org/10.1001/archpsyc.1982.04290080001001.

Liddle, P.F., Ngan, E.T.C., Caissie, S.L., Anderson, C.M., Bates, A.T., Quested, D.J., White, R., Weg, R., 2002. Thought and language index: an instrument for assessing thought and language in schizophrenia. Br. J. Psychiatry 181, 326–330. https://doi.org/10.1192/bjp.181.4.326.

Limisiewicz, T., Mareček, D., 2020. Syntax representation in word embeddings and neural networks—A survey. https://doi.org/10.48550/arXiv.2010.01063.

Lord, L.-D., Stevner, A.B., Deco, G., Kringelbach, M.L., 2017. Understanding principles of integration and segregation using whole-brain computational connectomics: implications for neuropsychiatric disorders. Philos. Trans. A Math. Phys. Eng. Sci. 375, 20160283 https://doi.org/10.1098/rsta.2016.0283.

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

Miller, T.J., McGlashan, T.H., Rosen, J.L., Cadenhead, K., Cannon, T., Ventura, J., McFarlane, W., Perkins, D.O., Pearlson, G.D., Woods, S.W., 2003. Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: predictive validity, interrater reliability, and training to reliability. Schizophr. Bull. 29, 703–715. https://doi.org/10.1093/oxfordjournals.schbul.a007040.

Montani, I., Honnibal, M., Honnibal, M., Landeghem, S.V., Boyd, A., Peters, H., McCann, P.O., geovedi, jim, O'Regan, J., Samsonov, M., Altinok, D., Orosz, G., Kok, D. de, Kristiansen, S.L., Bournhonesque, R., Miranda, L., Kannan, M., Baumgartner, P., Edward, Bot, E., Hudson, R., Roman, Fiedler, L., Howard, G., Phatthiyaphaibun, W., Tamura, Y., Mitsch, R., Bozek, S., murat, 2022. explosion/spaCy: v3.4.2: Latin and Luganda support, Python 3.11 wheels and more. https://doi.org/10.5281/zenodo.7228125.

Morgan, S.E., Diederen, K., Vértes, P.E., Ip, S.H.Y., Wang, B., Thompson, B., Demjaha, A., De Micheli, A., Oliver, D., Liakata, M., Fusar-Poli, P., Spencer, T.J., McGuire, P., 2021. Natural language processing markers in first episode psychosis and people at clinical high-risk. Transl. Psychiatry 11, 1–9. https://doi.org/10.1038/s41398-021-01722-y.

Murray, H.A., 1943. Thematic Apperception Test. Harvard University Press, Cambridge, MA, US.

Opler, M.G., Yang, L.H., Caleo, S., Alberti, P., 2007. Statistical validation of the criteria for symptom remission in schizophrenia: preliminary findings. BMC Psychiatry 7, 35. https://doi.org/10.1186/1471-244X-7-35.

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

Parola, A., Lin, J.M., Simonsen, A., Bliksted, V., Zhou, Y., Wang, H., Inoue, L., Koelkebeck, K., Fusaroli, R., 2022. Speech disturbances in schizophrenia: assessing cross-linguistic generalizability of NLP automated measures of coherence. Schizophr. Res. https://doi.org/10.1016/j.schres.2022.07.002.

Pasquiou, A., Lakretz, Y., Thirion, B., Pallier, C., 2023. Information-restricted neural language models reveal different brain regions' sensitivity to semantics. Syntax Context. https://doi.org/10.48550/arXiv.2302.14389.

Pauselli, L., Halpern, B., Cleary, S.D., Ku, B.S., Covington, M.A., Compton, M.T., 2018. Computational linguistic analysis applied to a semantic fluency task to measure derailment and tangentiality in schizophrenia. Psychiatry Res. 263, 74–79. https://doi.org/10.1016/j.psychres.2018.02.037.

Pei, Z., Li, C., Xiong, W., Luo, Z., 2020. Better modeling the hierarchical structure of language for sentence semantic matching. J. Phys.: Conf. Ser. 1651 (1), 012141 https://doi.org/10.1088/1742-6596/1651/1/012141.

Pintos, A.S., Hui, C.L.-M., De Deyne, S., Cheung, C., Ko, W.T., Nam, S.Y., Chan, S.K.-W, Chang, W.-C., Lee, E.H.-M., Lo, A.W.-F., Lo, T.-L., Elvevåg, B., Chen, E.Y.-H., 2022. A Longitudinal study of semantic networks in schizophrenia and other psychotic disorders using the word association task. Schizophr. Bull. Open 3, sgac054. https://doi.org/10.1093/schizbullopen/sgac054.

Pomarol-Clotet, E., Oh, T.M.S.S., Laws, K.R., McKenna, P.J., 2008. Semantic priming in schizophrenia: systematic review and meta-analysis. Br. J. Psychiatry 192, 92–97. https://doi.org/10.1192/bjp.bp.106.032102.

Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., 2021. Learning transferable visual models from natural language supervision. In: International Conference on Machine Learning. PMLR, pp. 8748–8763.

Reimers, N., Gurevych, I., 2019. Sentence-BERT: sentence embeddings using siamese BERT-networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Presented at the EMNLP-IJCNLP 2019. Association for Computational Linguistics, Hong Kong, China, pp. 3982–3992. https://doi.org/10.18653/v1/D19-1410.

Rochester, S., Martin, J.R., 1979. Crazy Talk: A Study of the Discourse of Schizophrenic Speakers. Springer, US.

Salazar, J., Liang, D., Nguyen, T.Q., Kirchhoff, K., 2020. Masked language model scoring. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Presented at the ACL 2020, Association for Computational Linguistics, pp. 2699–2712. https://doi.org/10.18653/v1/2020.acl-main.240. Online.

Sevilla, G., Rosselló, J., Salvador, R., Sarró, S., López-Araquistain, L., Pomarol-Clotet, E., Hinzen, W., 2018. Deficits in nominal reference identify thought disordered speech in a narrative production task. PLoS One 13, e0201545. https://doi.org/10.1371/journal.pone.0201545.

Silva, A.M., Limongi, R., MacKinley, M., Ford, S.D., Alonso-Sánchez, M.F., Palaniyappan, L., 2022. Syntactic complexity of spoken language in the diagnosis of schizophrenia: a probabilistic Bayes network model. Schizophr. Res. https://doi.org/10.1016/j.schres.2022.06.011. S0920-9964(22)00245–6.

Sun, X., Meng, Y., Ao, X., Wu, F., Zhang, T., Li, J., Fan, C., 2022. Sentence similarity based on contexts. Trans. Assoc. Comput. Linguist. 10, 573–588. https://doi.org/10.1162/tacl_a_00477.

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

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

Yi, E., Koenig, J.-P., Roland, D., 2019. Semantic similarity to high-frequency verbs affects syntactic frame selection. Cogn. Linguist. 30, 601–628. https://doi.org/10.1515/cog-2018-0029.