

Publication status: This preprint has been published elsewhere.

DOI of the published preprint: <https://doi.org/10.3389/frai.2026.1781552>

On the interface between Linguistics, Computer Science and Psychiatry: analyzing textual key-factors affecting BERT-based classification of schizophrenia in social media texts

João Victor Miranda e Silva, Cilene Rodrigues, Emilio Ashton Vital Brazil

<https://doi.org/10.1590/SciELOPreprints.14646>

Submitted on: 2025-12-18

Posted on: 2025-12-31 (version 1)

(YYYY-MM-DD)

On the interface between Linguistics, Computer Science and Psychiatry: analyzing textual key-factors affecting BERT-based classification of schizophrenia in social media texts

MIRANDA E SILVA, J.V.

Programa de Pós-Graduação em Estudos da Linguagem, PUC-Rio, Rio de Janeiro, Brazil.

<https://orcid.org/0000-0002-0525-5307>

RODRIGUES, C.

Programa de Pós-Graduação em Estudos da Linguagem, PUC-Rio, Rio de Janeiro, Brazil. IMPA Tech, Rio de Janeiro, Brazil.

<https://orcid.org/0000-0001-5324-7486>

VITAL BRAZIL, E. A.

IMPA Tech, Rio de Janeiro, Brazil.

<https://orcid.org/0000-0003-4982-3836>

*** Correspondence:**

João Victor Miranda e Silva

jvomiranda@gmail.com

Keywords: schizophrenia₁, language₂, data filtering₃, natural language processing₄.

Abstract

This paper investigates language impairments in schizophrenia (SZ) by integrating insights from language-centered investigations with computational approaches. Using BERT-base-cased, a transformer-based model, it explores how linguistic markers of SZ can be identified through Natural Language Processing (NLP) techniques, with emphasis on improving performance reliability via dataset refinement and approaching interpretability of deep learning outputs via statistical analyses of thematic content. We report the fine-tuning of a BERT model for text classification of 31,278 Reddit posts (15,639 SZ, 15,639 controls). The experiment evaluated the capacity of the model to distinguish language produced by individuals with SZ. The model achieved moderate performance (Accuracy = 0.6969; AUC = 0.78) and remained stable across hyperparameter configurations, indicating that foundation models such as BERT easily fit to data and, therefore, further performance gains are more likely to be derived from dataset refinement than from additional hyperparameter optimization. There were three key factors affecting the model's performance: text length, topic of discussion and vocabulary choices. Posts that were correctly classified tended to be significantly longer ($p < 0.001$, $M = 37.30$), focused on specific abstract topics (e.g., religion), and contained more words related to mental conditions. These factors have also been reported in manual analyses of the impacts of SZ on language. These findings contribute to the accuracy of computational models aimed at working on linguistic classification tasks and underscore the value of carefully curated datasets, while demonstrating the viability of NLP methods in profiling SZ language.

1 Introduction

Schizophrenia (SZ) is a mental disorder that affects cognitive and communicative abilities. It is characterized by disturbances in thought processes, emotional regulation, and social interactions,

often leading to changes in the perception of reality (APA, 2013). Among its cognitive impairments, language deficits stand out as a key area of investigation, given that individuals with SZ often exhibit difficulties in organizing their thoughts coherently, maintaining logical discourse, and adhering to combinatorial rules of language, including those responsible for structure building and constraints on grammatical and semantic dependencies. The presence of linguistic anomalies in SZ not only serves as a core diagnostic marker but also provides a unique perspective on the relation between language and cognition (Andreasen 1979a; Hinzen, 2017).

Given the centrality of language deficits in SZ, there is an increasing need to develop methodologies capable of systematically characterizing linguistic impairments associated with the disorder. In this scenario, not only methodologies derived from linguistics and cognitive sciences have positively contributed to a greater understanding of linguistic deficits attributed to SZ, but also Natural Language Processing (NLP) techniques have enabled large-scale analyses of SZ-related speech patterns across different linguistic domains, offering quantifiable measures of linguistic impairments (e.g., Birnbaum et al., 2017; Rezaii et al., 2019, among others). Studies leveraging language quantification methodologies have reported successful performances at classifying SZ through language use via systematic deviations from neurotypical speech. These reports are aligned with the SZ linguistic phenomena reported by investigations that apply specialized human annotated schemes (see for example Ziv et al., 2022; Guerra, 2023).

Moreover, NLP approaches to SZ have expanded beyond clinical data, increasingly incorporating social media sources due to the vast volume of linguistic samples available on such platforms (Mcmanus et al., 2015; Birnbaum et al., 2017; Guerra, 2023). While clinical data remains the gold standard, social media datasets provide the amount of data necessary for training large language models (LLMs), which are currently the state-of-the-art architecture for NLP tasks. However, despite their advantages in volume, social media datasets present challenges related to reliability and interpretability. The literature has shown, for instance, that classification models easily rely on lexical cues for classification (Guerra, 2023). Furthermore, although LLMs achieve state-of-the-art performance in NLP tasks and excel when fine-tuned on large datasets, general understanding of their underlying decision-making processes is still limited, reducing their applicability in clinical research (Carvalho et al., 2019; Bommasani et al., 2021).

Therefore, there is still a fundamental gap: lack of empirical understanding of which textual factors most strongly influence transformer-based classifications of SZ. That is, to what extent these models rely on linguistic structures, lexical cues, or topic patterns. Most studies emphasize performance metrics, but few systematically analyze the model's decision-making process or evaluate how dataset curation shapes interpretability and reliability, limiting contributions of NLP to theoretical linguistics and psychiatry. Against this backdrop, the current investigation addresses this gap by conducting an exploratory analysis of key textual factors that influence a BERT-based classification of SZ and control (non-SZ, NSZ) posts from Reddit. In conducting this research, we seek to: (i) fine-tune a transformer model for SZ classification with a carefully curated dataset designed to minimize lexical and topic bias, (ii) extract data-intrinsic information (text length, SZ-related lexical content, and discussion topic) that can influence data collection and processing in SZ-language studies.

Our findings show that a state-of-the-art transformer-based model can distinguish between SZ and NSZ posts at above-chance levels (~50%). However, the analysis indicates that the model relies on three data-intrinsic properties: text length, topic of discussion, and vocabulary choices. Posts that were correctly classified tended to be significantly longer ($p < 0.001$, $M = 37.30$), on specific abstract topics (e.g., religion) and contained more or less words related to mental conditions, with more use of

words semantically related to SZ triggering classification as SZ. This highlights the central role of dataset curation in ensuring the reliability and interpretability of computational models in SZ classification.

The paper is structured as follows. Section 2 presents an overview of how SZ affects language and reviews the main contributions of computational approaches to the study of SZ-related linguistic phenomena. Sections 3 to 5 detail our empirical study, describing the data collection and statistical methodology, reporting the results and discussing their implications to our understanding of how SZ affects language.

2.1 Language-grammar in face of schizophrenia

SZ is characterized by a spectrum of symptoms typically categorized as positive (e.g., hallucinations, delusions, disorganized speech), negative (e.g., alogia, anhedonia, affective flattening), and disorganized (e.g., deficits in executive function, memory, and verbal fluency) (APA, 2013). These symptoms broadly affect cognition, manifesting with varying degrees of severity and diverse combinations, which contribute to the complexity of SZ diagnosis. Importantly, linguistic anomalies emerge as some of the most direct manifestations of the disorder's underlying cognitive disruptions. Particularly, such dysfunctions are often associated with disorganized speech, being closely linked to the broader construct of Formal Thought Disorder (FTD). FTD is traditionally divided into positive (+FTD) and negative (-FTD) manifestations. While -FTD is associated with impoverished speech, such as a significant reduction in verbal output and meaning (poverty of speech and poverty of content), +FTD is linked to discourse disorganization, including derailment, tangentiality, incoherence, and illogicality (Andreasen and Grove, 1986; Mckenna and Oh, 2005).

Linguistic research over the past decades has shown that, particularly, +FTD correlates with a higher frequency of linguistic anomalies. Speakers with SZ +FTD frequently exhibit reduced syntactic complexity, impaired referential coherence, and fragmented phrasal and structures (Docherty et al., 1996; Çokal et al., 2018). Early work by Rochester and Martin (1979), for example, revealed a higher incidence of errors involving pronominal reference among SZ +FTD, while SZ -FTD tended to overuse full nominal expressions. More recent studies have corroborated these findings: Çokal et al. (2018) and Sevilla et al. (2018), for example, reported significantly higher rates of referential anomalies (e.g., more than one possible referent, no possible referent), specially in 3Person pronouns, among SZ +FTD participants, and Tovar Torres et al. (2019) observed that SZ +FTD is associated with a greater frequency of anomalous constructions inside Determiner Phrases (DPs).

Beyond referentiality, deficits appear in several grammatical domains. Research on morphology indicates that SZ speakers show difficulty with combinatorial morphology (forming the regular past tense), but not in accessing irregular lexical forms (Walenski et al., 2010). Studies on sentence production reveal reduced use of embedded clauses (Morice and McNicol, 1986; DeLisi, 2001; Çokal et al., 2018, Chaves, 2022) and increased rates of ungrammatical ellipsis (Chaves, 2022). Additionally, reduced use of determiners and preference for the use of null pronominal forms (Chaves, 2022; Chaves et al., 2023; Çokal et al., 2023), which are arguably simpler structures (Cardinaletti and Starke, 1994), are reported in SZ. These findings align with theoretical proposals that interpret SZ-related linguistic deficits, particularly +FTD, as disruptions in the combinatorial architecture of grammar itself (Hinzen, 2017). From this perspective, thought disorder reflects grammatical impairments, failures in the combinatorial processes underlying sentence production and comprehension (Moro et al., 2015) that lead to disordered thoughts.

Much of this evidence comes from studies that employ controlled data collection methodologies. Linguistic samples are typically elicited through standardized interviews or narrative tasks, which are aimed at ensuring consistency across speakers and allowing for more reliable comparison between SZ and NSZ (Çokal et al., 2018). Importantly, the literature emphasizes that linguistic disorganization is more readily observed in extended discourse. Andreasen (1979a, 1979b) showed that longer utterances increase the likelihood of observing +FTD linguistic patterns, and Docherty et al. (1996, 1997) required at least ten minutes of speech to capture communicative failures. In accordance, Chaves (2022) found that 30-second narratives (average 37.6 words) were insufficient to reveal SZ-specific grammatical markers. These observations are rather expected: longer discourse production naturally results in a larger amount of information, increasing, thus, statistical measures of grammatical structure (e.g., means and ratios).

In summary, linguistic impairments observed in SZ are not merely secondary symptoms of broader deficits. Language is arguably the main epicenter of the cognitive disruptions observed in SZ. Research evidence supports the notion that SZ breaks down syntactic structure, leading to semantic and pragmatic shortfalls. As language is conceptualized as a biological and cognitive self-contained system unique to humans, interfacing with various mental abilities (Hauser et al., 2002; Chomsky, 2006), understanding how SZ-related linguistic abnormalities emerge requires reliable and robust methodologies to assess language and can provide valuable insights into biomarkers of the disorder.

2.2 Schizophrenia and Natural Language Processing (NLP)

The advent of computational methods has expanded the scope of linguistic analyses. NLP is described as an ever-growing range of computational techniques aimed at automatically analyzing and representing human language (Cambria and White, 2014). For example, NLP approaches to investigating language can automatically identify grammatical categories and semantic relations, measuring similarity between words based on their contexts of occurrence, among other techniques.

In SZ studies, computational approaches have been used to automate the identification of linguistic anomalies, providing quantifiable measures of language deficits via large-scale investigations that surpass the capabilities of traditional linguistic analysis (Birnbaum et al., 2017; Rezaii et al., 2019; Ziv et al., 2022; Guerra, 2023). The results reported in the literature are promising, documenting linguistic patterns consistent with earlier clinical and theoretical analyses.

Ziv et al. (2022), for example, compared the transcribed oral narratives produced by SZ inpatients and NSZ via a morphological and a part-of-speech (POS) tagger to identify the lemma of each word, tense, number, gender and person features. Their results indicate a lower rate of inflectional morphology in SZ, alongside increased use of 1Person pronouns and decreased use 3Person forms. Such patterns are aligned with observations in formal linguistic studies. Birnbaum et al.'s (2017) and Bae et al.'s (2021), further corroborated observations in the pronominal domain by identifying, also via POS taggers, a greater volume of 1Person pronoun usage and variability in 3Person pronouns. Tang et al. (2021) further observed that SZ productions are characterized by fewer determiners.

Beyond automated annotation, Mota et al. (2014) analyzed narratives produced by formally diagnosed SZ patients (vs. NSZ) using speech graph analysis (SGA). Participants were asked to recount a recent dream and narrate waking daily routine activities. Their responses were transcribed and transformed into speech graphs, where words functioned as nodes and the temporal links between them as edges, creating a network that reflected the conceptual structure of speech. The analysis revealed that SZ produced significantly impoverished graphs characterized by fewer edges

and lower connectivity. This indicates reduced conceptual coherence and a fragmented semantic structure. These outcomes parallel observations of Chaves (2022) and Chaves et al. (2023), who noted SZ impairments particularly at the sentential level, including sentence embedding, coordination, and in the correct application ellipsis.

All together, these results indicate that clinical/manual and automated analyses of language in SZ could work hand in hand being used in different scientific settings: manual analyses are better suited for small, controlled datasets, while computational approaches are more effective for processing large, less controlled datasets, overcoming human limitations of scale and time that often restrict large-scale linguistic research. Also, while manual analyses provide a fine-grained in-depth description of the deficits observed in SZ, computational models could offer coarse-grained overview of these deficits, orienting, thus, manual analysis.

However, it's important to point out that, to achieve a desirable state of collaboration, automated analyses should work on the reliability of their results, considering the quality and intrinsic aspects of analyzed data. Evidence suggests that machine-learning models trained on social media data may rely heavily on surface-level cues. Mota et al. (2014) indicates that certain discussion topics, particularly those involving abstract thinking, are more sensitive to underlying linguistic and cognitive dysfunctions, a pattern consistent with results from Pugh et al. (2024), who shows that LLaMA 3 performed better when classifying fictional narratives. Also, according to Kayi et al. (2018), when classifying social media text, an algorithm can rely more on lexical choices (e.g., words indicating negative sentiment), while McManus et al. (2015) found that words indirectly associated with psychosis were disproportionately weighted by models.

Also, while machine learning algorithms trained on linguistic features have reported high classification accuracies for distinguishing SZ from control groups, typically ranging between 80% and 90% accuracy rate (Birnbaum et al., 2017; Bae et al., 2021; Ziv et al., 2022), computer models tend to overfit to strong contextual lexical cues. This concern is addressed by Guerra (2023), who attempted to reduce lexical bias by removing posts from r/schizophrenia and imposing a five-word minimum length, still obtaining an F-score above .90. In addition, as noted in Section 1, extremely short texts are insufficient for analyzing structural linguistic properties, limiting the interpretability of such results. Additionally, the r/schizophrenia subreddit is only one of many forums where discussions frequently revolve around mental health.

These concerns exemplify the so-called “black box” concern (Bommasani et al., 2021), which refers to the difficulty of interpreting how deep learning models process information through multiple complex layers of representation. These internal transformations often yield outputs that, while accurate, lack transparency regarding underlying decision-making mechanisms. Consequently, it's often non-trivial to identify which linguistic features are being more influential to a deep learning model's prediction. Since deep learning architectures are currently state-of-the-art in NLP approaches, it benefits the investigation of language in SZ to make use of such architectures to identify linguistic patterns related to the disorder. However, the interpretability of the relation between input features and output classifications is important for research fields such as psychiatry and linguistics, which are concerned with clear clues that can further expand the understanding of the relation between the disorder and its symptoms, ultimately leading to a possible improvement of diagnostic methods through language and comprehension of the architecture of Grammar and how it can be affected in neuroatypical minds. For instance, the reliance of these models on lexical SZ-related cues, as reported in several of the aforementioned studies, is less generalizable than

grammatical structure and therefore less effective for both diagnostic applications and theoretical linguistic analysis. Therefore, it's important to understand to which degree LLMs may rely on them.

Provoked by these concerns, and aiming at contributing to the dialogue between psychiatry, computer science and linguistics, we conducted an exploratory study to verify whether or not text-length, lexical bias and topic of discussion affect a BERT-based classification of SZ.

3 Materials and methods

3.1 Data collection process description

The dataset was collected from Reddit, comprising posts and comments from SZ and control users. Since mental health data are sensitive information, all samples were anonymized.

Users composing the SZ group were identified through participation in r/schizophrenia and self-identification via flairs, which are customizable tags that users can select to represent their profile within a subreddit (e.g., "schizophrenic"). Self-identification was further validated through regular expression (regex) searches for explicit self-declarations (e.g., "I am schizophrenic"), followed by manual verification. Additional filtering was applied to exclude users with conflicting self-reported mental disorders and non-English speakers using regex searches. A final filtering step removed subreddit-specific biases, excluding those focused on mental health (e.g., r/TalkTherapy, r/MentalHealthBabies) and non-English topics (e.g., r/askfrance, r/learndutch). A single mental health subreddit (r/AskDocs) was left in our dataset to represent SZ-related content for further investigation. Posts with fewer than five words were removed, and text normalization was applied using Python's unicodedata library (NFKD normalization, ASCII encoding). Links were removed. Stop words and punctuation were preserved. Duplicates were excluded. Control (NSZ) users were selected based on the top 10 subreddits with the highest post counts in the SZ group after all filtering was applied. We retrieved up to 100 unique users per subreddit using PRAW's top post queries with no overlap with SZ users. The same filtering criteria (regex searches for conflicting disorders, non-English speakers, and SZ declarations) was applied to NSZ data collection, including encoding normalization, and text-length thresholding. All posts from the NSZ group were derived from subreddits that are also present in the SZ set of posts. This was performed as a way of achieving discussion topics consistency across groups.

To mitigate subreddit distribution imbalance and reduce noise, another screening was conducted excluding subreddits with fewer than 10 unique contributors. NSZ group was composed by selecting posts from the same subreddits retained in the SZ group, using a one-to-one matching strategy to balance post counts between groups. When a given subreddit lacked enough NSZ posts to match its SZ counterpart, we supplemented the insufficiency with additional posts from the top 10 most represented SZ subreddits to preserve content consistency, aligning with the sufficient scale for BERT (DEVILIN et al., 2018). See the supplementary material available at the public data repository for a selected content flowchart. See Tables 1 and 2 for dataset descriptions.

Table 1. Data summary after content balance.

Data summary after content balance			
Content data	Groups		
	Target (SZ)	Control (NSZ)	Total sample

Number of posts	15,639	15,639	31,278
Total number of words	613,891	480,616	1,094,507
Total unique subreddits	114	114	114
Mean word count per post	39.25	30.73	34.99
Unique users	183	428	611

Table 2. Content summary of dataset after content balance.

Top 10 subreddits after content balance							
Target (SZ)				Control (NSZ)			
Subreddit	Posts	Mean word count	% posts of total	Subreddit	Posts	Mean word count	% posts of total
AskReddit	3740	32.65	24.33	AskReddit	4268	25.03	27.77
teenagers	670	19.56	4.36	teenagers	970	16.67	6.31
NoStupidQuestions	437	45.91	2.84	pics	564	24.42	3.67
pics	404	19.92	2.63	NoStupidQuestions	532	34.28	3.46
AskWomen	337	64.82	2.19	AskWomen	463	48.69	3.01
trees	308	29.41	2.00	memes	391	12.74	2.54
memes	302	20.11	1.96	facepalm	278	28.41	1.81
ftm	297	41.68	1.93	news	238	37.8	1.55
facepalm	262	27.88	1.70	AmItheAsshole	235	60.69	1.53
amiugly	251	30.33	1.63	unpopularopinion	222	39.79	1.44
Others	8631	44.76	56.16	Others	7478	34.51	48.66

The study in its totality was submitted to and approved by the PUC-Rio (Pontifical Catholic University of Rio de Janeiro) Research Ethics Committee under protocol number 78-2024, approval reference 91-2024.

3.2 Model hyperparameter optimization for fine-tuning

We fine-tuned a BERT-base-case model for a classification task to distinguish between SZ and NSZ posts from the dataset described above. This model was selected due to its efficient processing and its ability to preserve casing information, which can act as insightful features in social media texts (Herring, 2019). The model was fine-tuned using a supervised learning approach. The training data was split into 80% for training and 20% for testing. Fine-tuning was performed on labeled Reddit posts using a classification head on top of BERT's final hidden states. Given BERT's token limit of

512 per input, longer posts were truncated, while shorter ones were padded with special tokens. In total, the training dataset contained 1,174,764 non-padding tokens. The test dataset contained 283,587 non-padding tokens.

To analyze the performance consistency of the model and identify an optimal set of hyperparameters for the present classification task, we employed Optuna, an advanced hyperparameter optimization framework. Optuna utilizes a Bayesian optimization approach with Tree-structured Parzen Estimators, allowing for an efficient search over a high-dimensional space of hyperparameters (Akiba et al., 2019). In our implementation, we optimized the learning rate within a log-uniform range from $1e-5$ to $5e-5$ and the number of training epochs as an integer between 2 and 5. In addition to these optimized parameters, we maintained a set of immutable hyperparameters throughout the fine-tuning process, including a per-device training and evaluation batch size of 32 and a weight decay of 0.01. The model was fine-tuned using the Hugging Face Transformers library via a Python notebook, with all other hyperparameters kept at default settings.

For each trial, Optuna selected a new set of hyperparameters and evaluated their performance based on accuracy. The study spanned 15 trials.

The fine-tuning process was conducted using Google Colab Pro with an NVIDIA A100 GPU environment. The training scripts were implemented using the Hugging Face Transformers library and PyTorch. Performance metrics were extracted using the scikit-learn (sklearn) library.

3.3 Exploratory analyses of the model's outputs

Given the goals of our study, we conducted the following analyses:

- (a) Text length and prediction accuracy: We tested whether mean text length (extracted from the total word count of posts) significantly differed in correctly predicted posts. A one-way ANOVA was conducted with text length as the dependent variable and prediction as the independent variable.
- (b) SZ-related lexical bias: Posts from the r/AskDocs subreddit were manually inspected for SZ-related lexical items. Odds Ratio tests of independence assessed whether the presence of SZ-related terms was associated with correct model predictions.
- (c) Discussion topics and prediction accuracy: Chi-Squared tests of independence were applied to examine the association between subreddit membership and prediction outcomes. Subsequently, a logistic regression model was fit using r/AskReddit (the most represented subreddit) as the reference category to test whether other subreddits were significantly associated with correct model predictions.

Analyses of (a) and (b) were conducted in JASP (version 0.19.0.0). Logistic regression modeling was implemented in Google Colab using the libraries pandas, numpy and statsmodels. Visualizations were generated using matplotlib and seaborn. Datasets and Python notebooks are publicly available in the project's repository.

4 Results

4.1 Optuna hyperparameter study and best model's performance

The 15 Optuna trials were assessed based on validation accuracy, with the best model achieving an accuracy of 0.7022. The table below presents the details of each trial, including training loss at different steps and final validation accuracy. Note that trials that chose a number of epochs lower than 3 did not reach 2000 training steps. After evaluating all trials, the best-performing combination of hyperparameters regarding accuracy were learning rate fixed at 1.41e-5 and number of epochs at 3. See Table 3 for results.

Table 3. Results of the hyperparameters Optuna study. Best hyperparameter configuration highlighted. “LR” stands for “Learning Rate”, while Losses represent sets of 500 training steps.

Trial	LR	Epochs	1st Loss	2nd Loss	3rd Loss	4th Loss	Accuracy
1	1.25e-5	3	0.6152	0.5705	0.5376	0.4775	0.7011
2	1.11e-5	3	0.6183	0.5751	0.5446	0.4933	0.7019
3	1.98e-5	3	0.6183	0.5681	0.5249	0.4278	0.6963
4	3.36e-5	2	0.6204	0.5646	0.5105	-	0.7000
5	2.93e-5	2	0.6143	0.5580	0.5068	-	0.6973
6	3.64e-5	3	0.6206	0.5631	0.5136	0.3524	0.6870
7	2.79e-5	4	0.6221	0.5713	0.5203	0.3656	0.6762
8	1.76e-5	2	0.6168	0.5714	0.5342	-	0.7006
9	3.85e-5	3	0.6216	0.5645	0.5135	0.3387	0.6947
10	2.48e-5	5	0.6188	0.5688	0.5289	0.3852	0.6845
11	1.07e-5	5	0.6169	0.5731	0.5426	0.4768	0.6893
12	1.01e-5	4	0.6175	0.5745	0.5450	0.4882	0.6931
13	1.41e-5	3	0.6153	0.5717	0.5379	0.4640	0.7022
14	1.44e-5	4	0.6168	0.5704	0.5338	0.4500	0.6971
15	1.57e-5	3	0.6153	0.5704	0.5327	0.4485	0.7006

Once the optimal hyperparameters were identified, a new model was fine-tuned from scratch using the best hyperparameter configuration. Results are presented below.

Training: Loss decreased from 0.6149 (step 500) to 0.4571 (step 2000). Evaluation: Loss = 0.5933, Accuracy = 0.697, Precision = 0.686, Recall = 0.717, F1 = 0.701. TP/FP/TN/FN = 2224/1019/2136/877; Runtime = 36.8 s.

To further assess the model’s performance, a Receiver Operating Characteristic (ROC) curve is displayed (Figure 1), resulting in 0.78 AUC.

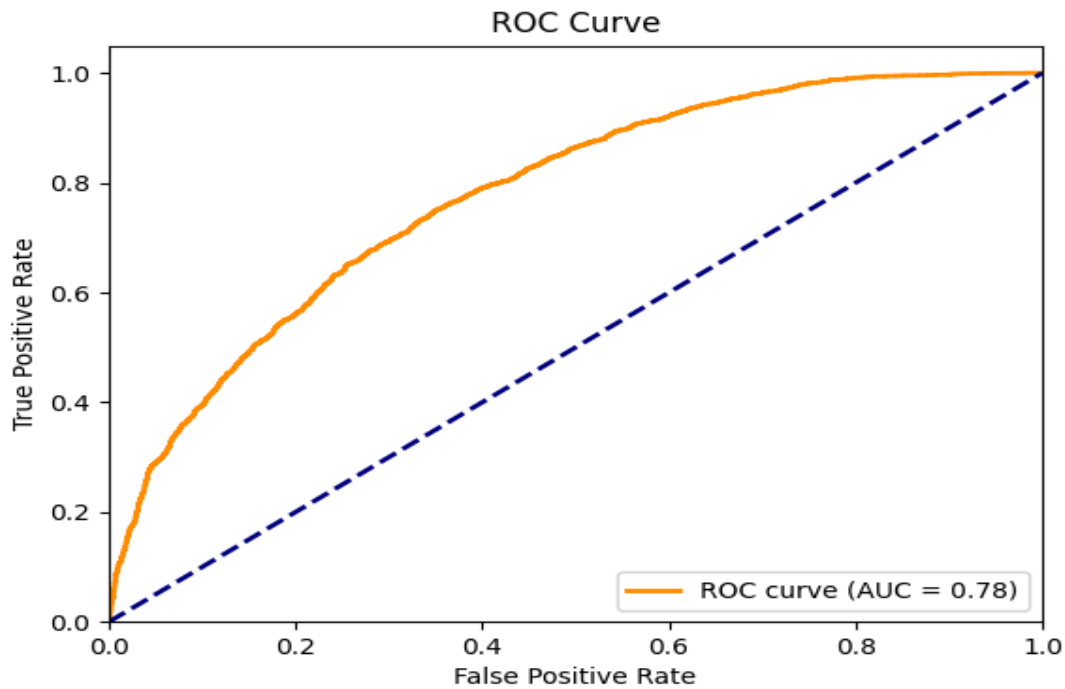


Figure 1. ROC Curve plot of the model’s test performance.

4.2 Impact of text length

Figure 2, showing the mean text length for each prediction class (TP, FP, TN, FN), indicates a notable difference between correct prediction categories (TN, TP) and incorrect ones (FN, FP), with the correct predictions exhibiting a higher mean text length.

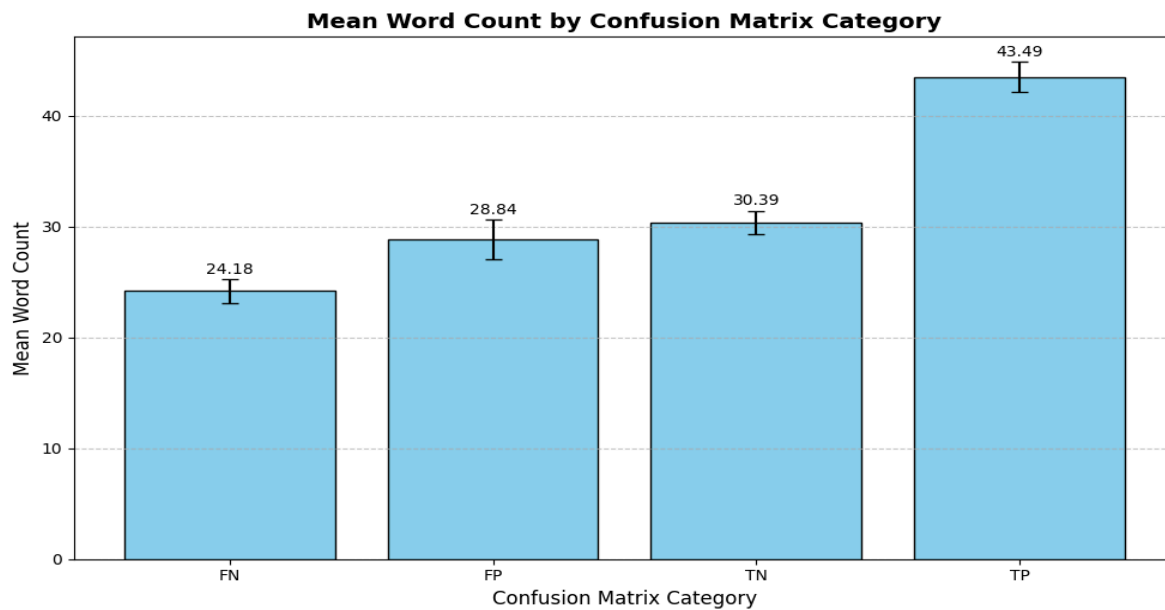


Figure 2. Bar graph with standard error of Mean Word Count (text length) Distribution by Confusion Matrix Category.

The results of the one-way ANOVA for text length across the four evaluation categories (TP, TN, FP, FN) reveal significant differences between correct and incorrect predictions (Table 4). There were significant differences in text length across categories, with an F-value of 38.486 and a p-value of less than 0.001. Post hoc comparisons revealed significant differences between TP and the other categories, particularly with FN, where the mean difference was the highest (19.312). FP and FN did not show a significant difference ($p = 0.245$), so as TN and FP ($p = 0.876$). Since text length (word count) distributions were non-normal, Kruskal–Wallis tests were conducted as a non-parametric robustness check. The test showed the same significance pattern as the ANOVA test ($p < 0.001$), supporting results. The effect size for the ANOVA was small ($\eta^2 = .018$).

Table 4. One-way ANOVA and post hoc comparisons of both the test dataset.

ANOVA Comparisons						
ANOVA - Text length						
Cases	Sum of Squares	df	Mean Square	F	p	η^2
Evaluation	339995.583	3	113331.861	38.486	< .001	.018
Residuals	$1.841 \times 10^{+7}$	6252	2944.721			
Post Hoc Comparisons - Evaluation						
		Mean Difference	SE	t	ptukey	
TP	TN	13.106	1.644	7.972	< .001	
	FP	14.658	2.053	7.140	< .001	
	FN	19.312	2.164	8.925	< .001	
TN	FP	1.552	2.066	0.751	0.876	
	FN	6.207	2.176	2.852	0.023	
FP	FN	4.655	2.500	1.862	0.245	

4.3 Lexical-cues bias

We tested r/AskDocs, a subreddit randomly selected to represent the broader theme of health. To evaluate effects of lexical bias, we manually reviewed all 18 samples from the subreddit and tagged posts containing mentions of psychosis, mental illness, or medications commonly prescribed for such conditions (e.g., clozapine). We then conducted Odds Ratio tests of independence to examine whether the presence of SZ-related terms was associated with correct model predictions.

A strong association between the presence of SZ-related words and the model's classification outcomes was found (Figure 3). Posts containing SZ-related terms were more likely to yield correct classifications (TP or TN). However, the confidence interval was wide and approached the line of no effect (red dashed line), reflecting the limited sample size and associated uncertainty. Posts mentioning SZ-related terms were significantly more likely to result in TP, representing the only association that clearly exceeded the no-effect threshold. In contrast, the Odds Ratio for TN was below 1, indicating no meaningful relationship and suggesting that such terms did not contribute to

correct non-SZ predictions. More details on the Odds Ratio tests of independence are available in the supplementary materials within the public data repository.

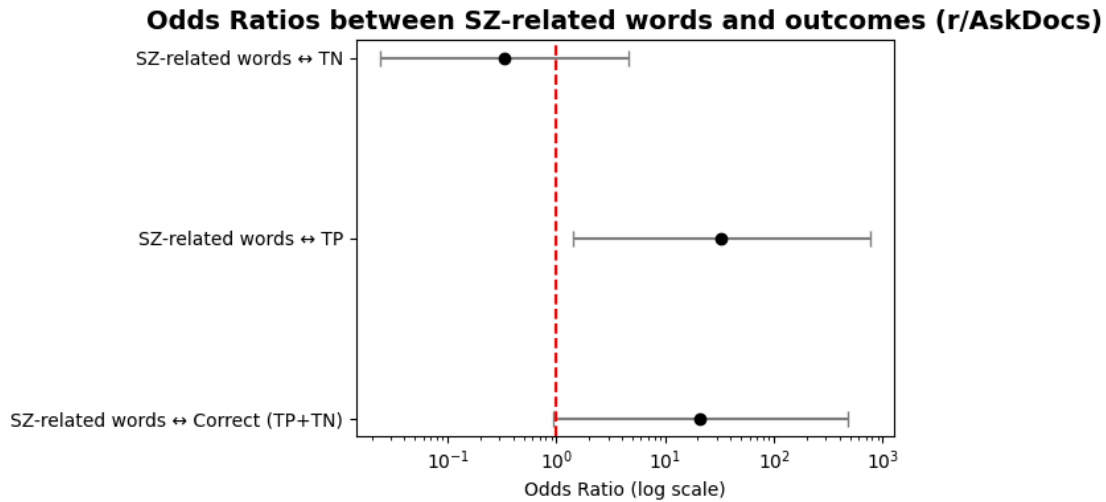


Figure 3. Odds Ratio (log scale) visual representation between SZ-related words and outcomes in r/AskDocs.

4.4 Topic of discussion

Chi-Squared tests of independence were conducted to examine the relationship between discussion topics and classification outcomes (i.e., TP, FP, TN, FN). Results revealed a statistically significant association between subreddits and classification outcomes ($X^2(113) = 194.680, p < .001$), with an effect size, measured by Cramer’s V, of 0.176, indicating a small to moderate association. To further explore which subreddits contribute most to correct predictions (TP, TN), we conducted an additional logistic regression analysis using subreddits as predictors. To examine whether certain subreddits were more likely to yield significant results, we set r/AskReddit as the reference category in the logistic regression model, as it was the most represented subreddit and provided a stable baseline for comparison. To address the large number of subreddits (114), we also estimated an L1-regularized logistic regression (LASSO). LASSO regularization performs variable selection by shrinking unstable coefficients. Still, all subreddits retained non-zero coefficients.

The model’s summary p-value of < 0.001 indicates that the influence of subreddits on correct predictions is statistically significant. The Pseudo-R² of 0.026 indicates that the model explains about 2.6% of the variance in log-likelihood compared to the null. Several subreddits were significantly associated with correct predictions, as indicated by the Wald test. Posts from r/Advice, r/AskWomen, r/Christianity, r/ftm, and r/suggestmeabook were significantly more likely to yield correct classification outcomes. In contrast, posts from r/CasualConversation, r/PublicFreakout, r/RoastMe, r/mildlyinfuriating, and r/pics were significantly less likely to result in correct predictions.

The results are summarized in Table 5, which shows the results of a logistic regression model examining the relationship between subreddit, inferred as discussion topics, and the likelihood of correct classification outcomes.

Table 5. Logistic regression model Summary - correct outcomes (TP, TN) as a function of subreddits.

Model Summary - correct outcomes (TP, TN) as a function of subreddits

Model	Deviance	AIC	BIC	df	ΔX^2	Pseudo R ²	p
$M_{\text{subreddits}}$	7477.6	7705.6	8475.4	6142	197. 8	0.026	< .001

Coefficients

Subreddit	Estimate	SE	z	df	p
Intercept*	0.590	0.184	3.211	1	0.001
r/Advice	0.908	0.347	2.615	1	0.009
r/AskWomen	0.464	0.198	2.339	1	0.019
r/CasualConversation	-1.095	0.424	-2.582	1	0.010
r/Christianity	1.401	0.433	3.235	1	0.001
r/PublicFreakout	-0.501	0.254	-1.976	1	0.048
r/RoastMe	-1.292	0.373	-3.467	1	0.001
r/ftm	0.820	0.368	2.229	1	0.026
r/mildlyinfuriating	-0.538	0.230	-2.340	1	0.019
r/pics	-0.388	0.150	-2.587	1	0.010
r/suggestmeabook	1.652	0.738	2.239	1	0.025

Estimates based on binomial categories (0,1). “0” represented incorrect classifications, while “1” represented correct classifications.

*Intercept is based on r/AskReddit, while the subreddits presented on this current Table are restricted to significant ones ($p < 0.05$) from the remaining 113 subreddits.

5 Discussion

The Optuna hyperparameter optimization yielded minimal variation across trials, with accuracy values clustering closely around 0.69-0.70. This confirms prior observations that transformer-based models fine-tune efficiently even with modest tuning and that performance differences within a narrow hyperparameter space are small when the dataset is sufficiently large (Devlin et al., 2018). The model’s final results are significant, as an accuracy near 70% and an above-chance AUC (.78) suggest it learned meaningful linguistic distinctions to discern classes rather than relied on random guessing. The relatively balanced trade-off between Precision (.686) and Recall (.717) indicates that the model is not biased toward either class, but rather identifies linguistic features that reliably differentiate SZ from NSZ in correct predictions. Furthermore, the consistent performance of the BERT-based model across multiple Optuna trials points to the conclusion that future efforts may be more impactful if directed towards improving dataset quality, thus addressing the classification challenge from a different axis.

Notably, our model’s performance is lower than that reported in studies utilizing less curated datasets. For instance, Guerra (2023) reported an F1-score above 0.91 using a similar BERT-based architecture. This discrepancy underscores the importance of dataset composition: when lexical or topical cues are overlooked during training or fine-tuning, models may report high scores but rely on features that are tangential to linguistic impairments associated with SZ. By contrast, stricter filtering,

as implemented here, seems to reduce this inflation, resulting in performance levels comparable to those from standardized clinical datasets (Kayi et al., 2018). This suggests that approximately 0.70 accuracy performance represents a baseline for models trained on data where lexical bias is minimized.

Our main observations supported that text length positively contributes to classification accuracy. Posts categorized as correct predictions (TP, TN) consistently exhibited higher word counts, while shorter posts yielded more misclassifications. This suggests that longer texts provide the model with greater contextual density, potentially including structural discourse markers (e.g., sentence embeddings, functional categories, etc.), more coherence relations, and even more lexical cues that facilitate prediction. However, the specific relationship between grammatical structure, coherence and the attention allocated by the model to such structures remains an open question not addressed in the present study.

These findings align with clinical-standardized linguistic research showing that SZ-related anomalies, particularly those associated with +FTD, are more observable in extended discourse (Andreasen, 1979a; Çokal et al., 2018; Chaves, 2022). Longer linguistic productions tend to expose cumulative disruptions in, for instance, coherence and structural organization. For computational models, longer inputs similarly increase the number of classification cues available. Consequently, minimum-length thresholds should be treated as methodological precautions in computational studies, helping ensure that analyses are based on texts that contain sufficient linguistic material for meaningful classification.

Also, as shown in 3.3, the performance metrics of models trained or fine-tuned on less rigorously filtered datasets may be inflated due to lexical biases. Our analysis of the r/AskDocs subreddit revealed that posts containing explicit SZ-related terms were significantly more likely to yield TP predictions.¹ Sample size was small (18) and, consequently, results should be interpreted as exploratory rather than conclusive. However, this pattern mirrors concerns raised in the literature about lexical overfitting in mental-health NLP models (Mcmanus et al., 2015; Kayi et al., 2018). Even after extensive dataset filtering, results indicate that the model still relied on explicit disorder-related vocabulary when available. This has implications on interpretability and generalizability of NLP models. Regarding the former, high performance in health-adjacent subreddits, or their overrepresentation in datasets, may reflect attention to lexical cues rather than grammatical structure. Models trained under such conditions risk misclassification in contexts where mental-health vocabulary is absent, limited, or culturally encoded differently. This impacts generalizability and is especially relevant for cross-linguistic research and for applications involving populations with varying literacy or discourse norms. Thus, we conclude that, while lexical cues are difficult to be entirely eliminated from datasets, their influence must be quantified and controlled to prevent misleading optimistic interpretations of classification performance.

Moreover, the logistic regression revealed that the inclusion of subreddits as predictors significantly improved model fit ($\Delta X^2 = 197.8$, $p < .001$). Text-oriented subreddits involving suggestion- or advice-giving or personal disclosures (e.g., r/Advice, r/AskWomen, r/Christianity, r/suggestmeabook) were associated with higher rates of correct predictions, whereas subreddits associated with lower accuracy were mostly visually oriented (e.g., r/pics, r/RoastMe). Even though the literature hypothesizes that topicality may impact the saliency of semantic coherence impairments (Mota et al., 2014), it's not yet possible to interpret the present results as evidence that BERT captures deeper

¹ Examples of posts classified as containing SZ-related terms are available as supplement material in our public repository.

grammatical or coherence disruption patterns related to topicality and SZ. Instead, topic differences likely influence the density of linguistic information available to the model (visual-oriented vs. text-heavy topics), the register or communicative style associated with different social media contexts (Herring, 2019) and the stability of discourse structure, which may vary across genres. For example, advice-oriented posts tend to include explicit temporal-causal sequences (e.g., “this happened, so I did X”, “this is the author who also wrote X”), whereas subreddits centered on images or jokes rely on short, informal replies with fewer linguistic contextual cues.²

This stresses that topic curation should, for now, be treated primarily as a methodological variable. Consequently, analyses that draw conclusions about SZ-related language should consider the interaction between linguistic features and discourse genre. However, one subreddit that was associated with a significantly negative impact on the model’s predictions, r/CasualConversation, was not image-centered and exhibited a mean length (42 words) comparable to the overall mean observed across correct prediction classes. This indicates that text length and topic do not appear to interact directly in shaping prediction outcomes. Rather, each seems to contribute additively, without strong interaction. This is supported by the text-length ANOVA, which indicates that prediction classes accounted for only ~2% ($\eta^2 = .018$) of the variance in post length. Further research is therefore needed to examine the linguistic patterns characteristic of subreddits positively associated with accurate classification and to compare them with those of r/CasualConversation. Investigating structure-oriented features (e.g., POS distributions, syntactic depth) alongside lexical patterns and/or semantic networks would help clarify which linguistic cues the model finds most informative and how these cues interact with dataset-intrinsic properties such as topic and length.

To conclude, we have shown that a BERT-based model achieves moderate success classifying posts/texts from SZ and NZS Reddit users. When it runs unconstrained it has a better performance. However, as shown, it used SZ-related words to do the required classification, impairing generalizability. Also, short posts induce more classification errors, whereas extensive posts lead to classificatory precision. While this study advanced our understanding of the relationship between data-intrinsic features and the computational assessment of SZ, some limitations must be acknowledged. First, it relies on self-declared diagnoses from Reddit users, which is in itself a noise in the dataset that can cause misclassifications from outset. We tried to reduce this noise filtering the data as discussed on section 3.1, but future work should improve this methodology.

Our investigation focused on revealing some textual properties that the model uses to classify a given text/post as SZ or NSZ. Thus, grammatical issues were not considered in this study. We acknowledge, however, that future studies should incorporate syntactic analyses to further our understanding on how SZ affects language and on how LLMs perform classification tasks of the sort discussed here.

6 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

7 Author Contributions

² See the materials in the public repository for examples of posts drawn from the relevant subreddits (e.g., r/pics, r/Advice).

J.V.M.S. collected and analyzed the data and was the primary author of the manuscript. C.R. advised the research, contributed to the interpretation of results, and participated in writing and revising the manuscript. E.A.V.B. coadvised the project, offering methodological support and critical feedback on data analysis and interpretation. All authors read and approved the final version of the manuscript and agree to be accountable for its content.

8 Funding

This research was funded by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil) through a PhD Fellowship assigned to the first author and a Research Productivity Grants (#312004/2022-2).

9 Data Availability Statement

Data collected and analyzed for this study can be found at:

<https://doi.org/10.5281/zenodo.17643919>

10 References

- Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework., in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, (Anchorage AK USA: ACM), 2623–2631. doi: 10.1145/3292500.3330701
- Andreasen, N. C. (1979a). Thought, language, and communication disorders: I. Clinical assessment, definition of terms, and evaluation of their reliability. *Arch. Gen. Psychiatry* 36, 1315–1321.
- Andreasen, N. C. (1979b). Thought, language, and communication disorders: II. Diagnostic significance. *Arch. Gen. Psychiatry* 36, 1325–1330.
- Andreasen, N. C., and Grove, W. M. (1986). Thought, language, and communication in schizophrenia: diagnosis and prognosis. *Schizophr. Bull.* 12, 348–359.
- APA (2013). *Diagnostic and statistical manual of mental disorders: DSM-5TM, 5th ed.* Arlington, VA, US: American Psychiatric Publishing, Inc. doi: 10.1176/appi.books.9780890425596
- Bae, Y. J., Shim, M., and Lee, W. H. (2021). Schizophrenia detection using machine learning approach from social media content. *Sensors* 21, 5924.
- Birnbaum, M. L., Ernala, S. K., Rizvi, A. F., De Choudhury, M., and Kane, J. M. (2017). A collaborative approach to identifying social media markers of schizophrenia by employing machine learning and clinical appraisals. *J. Med. Internet Res.* 19, e7956.
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., Arx, S. von, et al. (2022). On the Opportunities and Risks of Foundation Models. doi: 10.48550/arXiv.2108.07258
- Cambria, E., and White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Comput. Intell. Mag.* 9, 48–57.

- Cardinaletti, A., and Starke, M. (1995). The typology of structural deficiency: On the three grammatical classes. *ZAS Pap. Linguist.* 1, 1–55.
- Carvalho, D. V., Pereira, E. M., and Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics* 8, 832.
- Chaves, M. D. F. F. (2022). Structural deficiency in Schizophrenia: an exploratory study of the nominal and sentential domains. [dissertation/master's thesis]. [Rio de Janeiro (RJ)]: PUC-Rio.
- Chaves, M. F., Rodrigues, C., Ribeiro, S., Mota, N. B., and Copelli, M. (2023). Grammatical impairment in schizophrenia: An exploratory study of the pronominal and sentential domains. *PloS One* 18, e0291446.
- Chomsky, N. (2006). *Sobre natureza e linguagem*. Martins Fontes.
- Çokal, D., Palominos-Flores, C., Yalınçetin, B., Türe-Abacı, Ö., Bora, E., and Hinzen, W. (2023). Referential noun phrases distribute differently in Turkish speakers with schizophrenia. *Schizophr. Res.* 259, 104–110.
- Çokal, D., Sevilla, G., Jones, W. S., Zimmerer, V., Deamer, F., Douglas, M., et al. (2018). The language profile of formal thought disorder. *Npj Schizophr.* 4, 18. doi: 10.1038/s41537-018-0061-9
- DeLisi, L. E. (2001). Speech disorder in schizophrenia: review of the literature and exploration of its relation to the uniquely human capacity for language. *Schizophr. Bull.* 27, 481–496.
- Deneault, A., Dumais, A., Désilets, M., and Hudon, A. (2024). Natural language processing and schizophrenia: a scoping review of uses and challenges. *J. Pers. Med.* 14, 744.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding., in *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, 4171–4186. Available at: https://aclanthology.org/N19-1423/?utm_campaign=The+Batch&utm_source=hs_email&utm_medium=email&_hsenc=p2ANqtz-_m9bbH_7ECE1h3lZ3D61TYg52rKpifVNjL4fvJ85uqgg rXsWDBTB7YooFLJeNXHWqhvOyC (Accessed December 15, 2025).
- Docherty, N. M., DeRosa, M., and Andreasen, N. C. (1996). Communication disturbances in schizophrenia and mania. *Arch. Gen. Psychiatry* 53, 358–364.
- Docherty, N. M., Miller, T. N., and Lewis, M. A. (1997). Communication disturbances in the natural speech of schizophrenic patients and non-schizophrenic parents of Patients. *Acta Psychiatr. Scand.* 95, 500–507. doi: 10.1111/j.1600-0447.1997.tb10138.x
- Guerra, S. (2023). Contextual NLP Explanations for Language Biomarker Research: Identification of Schizophrenia Traits on Social Media Posts using Multilevel Part of Speech Feature. [dissertation/master's thesis]. [Minas Gerais (MG)]: UFMG.

Habtewold, T. D., Rodijk, L. H., Liemburg, E. J., Sidorenkov, G., Boezen, H. M., Bruggeman, R., et al. (2019). Implications of data-driven analyses for personalized therapy in psychosis: a systematic review of cluster-and trajectory-based modelling studies. Available at: <https://europepmc.org/article/ppr/ppr75508> (Accessed December 15, 2025).

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science* 298, 1569–1579. doi: 10.1126/science.298.5598.1569

Herring, S. C. (2012). Grammar and electronic communication. *Encycl. Appl. Linguist.*, 1–9.

Hinzen, W. (2017). Reference Across Pathologies: A New Linguistic Lens on Disorders of Thought. *Theor. Linguist.* 43. doi: 10.1515/tl-2017-0013

Huang, Y.-J., Lin, Y.-T., Liu, C.-C., Lee, L.-E., Hung, S.-H., Lo, J.-K., et al. (2022). Assessing schizophrenia patients through linguistic and acoustic features using deep learning techniques. *IEEE Trans. Neural Syst. Rehabil. Eng.* 30, 947–956.

Kayi, E. S., Diab, M., Pauselli, L., Compton, M., and Coppersmith, G. (2017). Predictive linguistic features of schizophrenia., in *proceedings of the 6th joint conference on lexical and computational semantics (* SEM 2017)*, 241–250. Available at: <https://aclanthology.org/S17-1028/> (Accessed December 15, 2025).

McKenna, P. J., and Oh, T. M. (2005). *Schizophrenic speech: Making sense of bathroofs and ponds that fall in doorways*. Cambridge University Press. Available at: <https://books.google.com/books?hl=pt-BR&lr=&id=4HBjybKPKQsIC&oi=fnd&pg=PA1&dq=MCKENNA,+Peter%3B+OH,+Tomasina.+Schizophrenic+speech:+Making+sense+of+bathroofs+and+ponds+that+fall+in+doorways.+Cambridge+University+Press,+2005.&ots=dmpg0IjUY2&sig=S-dWVyd95Mz7ipNKUqi-BF-6TYI> (Accessed December 15, 2025).

McManus, K., Mallory, E. K., Goldfeder, R. L., Haynes, W. A., and Tatum, J. D. (2015). Mining Twitter data to improve detection of schizophrenia. *AMIA Summits Transl. Sci. Proc.* 2015, 122.

Miranda e Silva, J. V., Chaves, M., and Rodrigues, C. (2024). Language in schizophrenia: interpreting structural deficits. *Revista Linguística* 20. doi: <https://doi.org/10.31513/linguistica.2024.v20n3a65561>

Morice, R., and McNicol, D. (1986). Language changes in schizophrenia: a limited replication. *Schizophr. Bull.* 12, 239–251.

Moro, A., Bambini, V., Bosia, M., Anselmetti, S., Riccaboni, R., Cappa, S. F., et al. (2015). Detecting syntactic and semantic anomalies in schizophrenia. *Neuropsychologia* 79, 147–157.

Mota, N. B., Furtado, R., Maia, P. P., Copelli, M., and Ribeiro, S. (2014). Graph analysis of dream reports is especially informative about psychosis. *Sci. Rep.* 4, 3691.

Pugh, S. L., Chandler, C., Cohen, A. S., Diaz-Asper, C., Elvevåg, B., and Foltz, P. W. (2024). Assessing dimensions of thought disorder with large language models: The tradeoff of accuracy and consistency. *Psychiatry Res.* 341, 116119.

Rezaii, N., Walker, E., and Wolff, P. (2019). A machine learning approach to predicting psychosis using semantic density and latent content analysis. *NPJ Schizophr.* 5, 9.

Rizzi, L., and Cinque, G. (2016). Functional Categories and Syntactic Theory. *Annu. Rev. Linguist.* 2, 139–163. doi: 10.1146/annurev-linguistics-011415-040827

Rochester, S., and Martin (2013). *Crazy talk: A study of the discourse of schizophrenic speakers*. Springer Science & Business Media. Available at: https://books.google.com/books?hl=pt-BR&lr=&id=uDSgBQAAQBAJ&oi=fnd&pg=PR3&dq=ROCHESTER,+Sherry%3B+MARTIN,+James.+Crazy+Talk:+A+Study+of+the+Discourse+of+Schizophrenic+Speakers.+New+York,+NY:+Plenum+Press,+1979.&ots=N3_pVwyrvh&sig=g9iL7jMEkLUDn9wRfOciGBVi3Kw (Accessed December 15, 2025).

Sevilla, G., Rosselló, J., Salvador, R., Sarró, S., López-Araquistain, L., Pomarol-Clotet, E., et al. (2018). Deficits in nominal reference identify thought disordered speech in a narrative production task | PLOS One. 13, e0201545. doi: <https://doi.org/10.1371/journal.pone.0201545>

Tang, S. X., Kriz, R., Cho, S., Park, S. J., Harowitz, J., Gur, R. E., et al. (2021). Natural language processing methods are sensitive to sub-clinical linguistic differences in schizophrenia spectrum disorders. *Npj Schizophr.* 7, 25.

Tovar Torres, A., Schmeisser Nieto, W. S., Gari Soler, A., Morey Matamalas, C., and Hinzen, W. (2019). Language disintegration under conditions of severe formal thought disorder. Available at: <https://repositori.upf.edu/items/9961401e-4a9c-48ff-83ef-29452b62f5ac> (Accessed December 15, 2025).

Walenski, M., Weickert, T. W., Maloof, C. J., and Ullman, M. T. (2010). Grammatical processing in schizophrenia: Evidence from morphology. *Neuropsychologia* 48, 262–269.

This preprint was submitted under the following conditions:

- The authors declare that the necessary Terms of Free and Informed Consent of participants or patients in the research were obtained and are described in the manuscript, when applicable.
- The authors declare that the preparation of the manuscript followed the ethical norms of scientific communication.
- The authors declare that they are aware that they are solely responsible for the content of the preprint and that the deposit in SciELO Preprints does not mean any commitment on the part of SciELO, except its preservation and dissemination.
- The authors declare that the data, applications, and other content underlying the manuscript are referenced.
- The deposited manuscript is in PDF format.
- The authors declare that the research that originated the manuscript followed good ethical practices and that the necessary approvals from research ethics committees, when applicable, are described in the manuscript.
- The authors declare that once a manuscript is posted on the SciELO Preprints server, it can only be taken down on request to the SciELO Preprints server Editorial Secretariat, who will post a retraction notice in its place.
- The authors agree that the approved manuscript will be made available under a [Creative Commons CC-BY](#) license.
- The submitting author declares that the contributions of all authors and conflict of interest statement are included explicitly and in specific sections of the manuscript.
- The authors declare that the manuscript was not deposited and/or previously made available on another preprint server or published by a journal.
- If the manuscript is being reviewed or being prepared for publishing but not yet published by a journal, the authors declare that they have received authorization from the journal to make this deposit.
- The submitting author declares that all authors of the manuscript agree with the submission to SciELO Preprints.