

Analysis of Verbal Fluency in Slovenian Language in Patients With Schizophrenia

Mila Marinković
mm9136@student.uni-lj.si
Faculty of Computer and Information Science,
University of Ljubljana
Večna pot 113
SI-1000 Ljubljana, Slovenia

Martina Zakšek
martina.zaksek@gmail.com
Splošna bolnišnica Celje
Oblakova ulica 5
SI-3000 Celje, Slovenia

Polona Rus Prelog
polona.rus@psih-klinika.si
Univerzitetna psihiatrična klinika Ljubljana
Chengdujska 45
SI-1000 Ljubljana, Slovenia

Jure Žabkar
jure.zabkar@fri.uni-lj.si
Faculty of Computer and Information Science,
University of Ljubljana
Večna pot 113, Ljubljana
SI-1260 Ljubljana, Slovenia

ABSTRACT

This study investigates verbal fluency in the Slovenian language among individuals diagnosed with schizophrenia compared to healthy controls. Participants completed a verbal fluency task, which involved producing as many words as possible starting with a specific letter in Slovenian within a set time limit. The analysis included statistical testing and semantic similarity measures using FastText embeddings. Significant differences were found between the groups in terms of the number of correct and total words produced. While semantic similarity showed minimal differences, global optimality divergence revealed notable disparities. These findings highlight the utility of comprehensive analytical approaches in understanding verbal fluency deficits in schizophrenia, emphasizing the need for nuanced methods to capture the complexity of cognitive impairments in this population.

KEYWORDS

Verbal fluency, Schizophrenia, Slovenian Language, Semantic Analysis, Statistical Analysis

1 INTRODUCTION

Verbal fluency tests are widely used to assess cognitive function and linguistic abilities in various clinical populations, including individuals with schizophrenia. These tests, which require participants to say words based on specific criteria, provide valuable insights into semantic memory, executive function, and language processing capabilities.

Schizophrenia is a chronic mental disorder characterized by symptoms such as cognitive disorganization, impaired semantic processing, and executive dysfunction. These symptoms often manifest as deficits in verbal fluency, where affected individuals typically produce fewer words and commit more errors, such as repetitions, intrusions, and neologisms. Understanding these verbal fluency deficits is crucial for developing targeted cognitive and linguistic interventions.

Previous studies [1–8] have documented that individuals with schizophrenia exhibit notable impairments in verbal fluency tasks,

producing fewer words and making more errors compared to healthy controls. To address this gap, we employed a comprehensive analytical approach combining traditional statistical tests with advanced semantic similarity measures using FastText embeddings.

We hypothesized that while local semantic relationships might not differ significantly between groups, broader semantic coherence and structural organization of speech would be markedly impaired in schizophrenia. By leveraging advanced techniques such as FastText embeddings, we aimed to uncover deeper insights into the semantic characteristics of verbal fluency in schizophrenia.

In this paper, we present a detailed analysis of verbal fluency performance in individuals with schizophrenia compared to healthy controls. We discuss the implications of our findings for understanding the cognitive and linguistic disruptions associated with schizophrenia and propose directions for future research to further explore these impairments.

2 RELATED WORK

The analysis of verbal fluency in individuals with schizophrenia has been extensively researched to understand the cognitive and neural mechanisms underlying the disorder. This study draws upon several key pieces of related work that have influenced our methodology and analytical approaches.

Nour et al. [8] investigated the semantic trajectories in schizophrenia by analyzing verbal fluency tasks using a computational model of word embeddings. Their study highlighted the reduced semantically guided word selection in people with schizophrenia and its correlation with hippocampal disruptions. This approach underscored the importance of semantic distance in understanding cognitive disorganization in schizophrenia and inspired our use of FastText to compute word embeddings and analyze semantic distances between words generated during verbal fluency tasks.

Galaverna et al. [1] conducted a detailed analysis of errors in verbal fluency tasks among individuals with chronic schizophrenia. Their research emphasized the prevalence of perseverative and intrusion errors in verbal fluency tasks, highlighting significant moderators such as the severity of negative symptoms, formal

thought disorder, and pharmacological variables. This study provided crucial insights into the patterns of errors (intrusions, repetitions, neologisms) in verbal fluency tasks, which are essential for understanding the cognitive deficits associated with schizophrenia.

Ojeda et al. [5] explored the relationship between verbal fluency and other cognitive domains in patients with schizophrenia and healthy controls. Their findings indicated that while healthy controls' verbal fluency was primarily predicted by processing speed, in patients with schizophrenia, it was more closely related to working memory. This study highlights the differing cognitive mechanisms underlying verbal fluency performance in schizophrenia and informed our consideration of different cognitive variables in our analysis.

Grimes et al. [2] examined the stability of verbal fluency abilities in outpatients with schizophrenia over a one-year period. They found that verbal fluency abilities remained stable over time, providing evidence against significant longitudinal decline in these cognitive domains among individuals with chronic schizophrenia. This study's findings on the stability of verbal fluency informed our understanding of the temporal consistency of cognitive impairments in schizophrenia.

Lehtinen et al. [4] presented a systematic administration and analysis approach for verbal fluency tasks, highlighting the importance of detailed scoring guidelines and exploring various underlying cognitive processes. Their method provided strong inter-rater reliability and demonstrated significant effects of education and gender on verbal fluency performance, reinforcing the need for comprehensive analysis beyond total scores. This study's emphasis on clustering, switching, and error analysis informed our analytical approach to understanding the cognitive processes involved in verbal fluency tasks in schizophrenia.

Kosmidis et al. [3] studied verbal fluency in institutionalized patients with schizophrenia, focusing on age-related performance decline. They found that elderly patients exhibited a disproportionate decline in phonemic fluency compared to younger patients, while semantic fluency remained relatively stable. This research highlighted the impact of aging on cognitive strategies like clustering and switching, which are critical for verbal fluency tasks. The findings underscore the importance of considering age and institutionalization duration when analyzing verbal fluency in schizophrenia.

Nogueira et al. [7] provided normative data on semantic and phonemic verbal fluency tasks for a European Portuguese population, considering the effects of age, gender, and education. Their study demonstrated that age and education significantly affect verbal fluency performance, while gender has a more variable impact. This research supports the need for demographic adjustments in verbal fluency assessments and informed our methodology in adjusting for these variables in our analysis.

These studies collectively emphasize the multifaceted nature of verbal fluency impairments in schizophrenia, necessitating the use of advanced analytical techniques to capture the underlying cognitive and linguistic disruptions. Our approach, combining traditional statistical tests with semantic similarity measures using FastText embeddings, aims to build on this foundation to provide deeper insights into the verbal fluency deficits in schizophrenia.

3 METHODOLOGY

This section outlines the methodology used in our study to investigate verbal fluency deficits in individuals with schizophrenia compared to healthy controls. We describe the participants, procedures, data collection, and data analysis methods employed to gather and analyze the data.

3.1 Participants

The study involved a total of 126 participants, divided into two groups: 58 individuals diagnosed with schizophrenia and 68 healthy controls. The participants were matched for age and gender to ensure comparability between the groups. All participants were 18 years or older. Exclusion criteria included an inability to speak Slovenian, a history of intellectual disability, organic brain conditions, or substance abuse. For healthy controls, additional criteria included no history of psychiatric disorders or substance abuse. The study was approved by the Medical Ethics Committee of the Republic of Slovenia, and all participants provided written informed consent.

3.2 Procedure

Participants were asked to perform a verbal fluency task in which they had to say as many words as possible that start with the letter "L" in Slovenian within one minute. This task was administered individually in a quiet room to minimize distractions. All responses were recorded for subsequent analysis. Demographic data, including age, gender, education level, marital status, employment status, and hospitalization history, were collected prior to the test to provide a comprehensive overview of the sample population.

3.3 Data Collection

All data were recorded and stored in a secure database. The verbal fluency responses were transcribed and annotated for analysis. Each word was evaluated for its accuracy and categorized as correct, intrusion (an incorrect word not fitting the criteria), repetition (same word used more than once), or neologism (made-up word). The timestamps for each word were also recorded to facilitate temporal analysis. Additionally, FastText embeddings were computed for each word to enable semantic similarity analysis.

Table 1: Demographic Characteristics of the Participants.

| Measure | Schizophrenia Patients | Healthy Controls |
|------------------------------------|------------------------|------------------|
| Total Participants | 58 | 68 |
| Average Age (years) | 46.05 | 46.71 |
| Prevalent Education Level | Primary school | High school |
| Avg. Elementary School grade-point | 3.57 | 4.72 |
| Avg. Secondary School grade-point | 3.46 | 4.39 |
| Male Distribution | 29 | 35 |
| Female Distribution | 29 | 33 |

Demographic data, summarized in Table 1, were analyzed to ensure that the groups were comparable in terms of age and gender, so that any differences observed in verbal fluency performance are less likely to be confounded by these factors. Although there were differences in the average education level between the schizophrenia

and healthy control groups, we verified that within each education level, there were no significant differences between the groups, ensuring that education level did not confound the verbal fluency comparisons.

3.4 Data Analysis

The collected data underwent various analyses to explore the differences in verbal fluency between individuals with schizophrenia and healthy participants. The following analytical techniques were employed:

- (1) **Statistical Analysis:** A t-test was conducted to compare the total number of words and the number of correct words produced by the two groups, where correct words are those that are neither intrusions, repetitions, nor neologisms. Before using the t-test, we checked that the data was normally distributed. This statistical test provided evidence of differences in verbal fluency performance between individuals with schizophrenia and healthy controls.
- (2) **Semantic Similarity Analysis:** For this study, we used FastText embeddings to capture semantic relationships between words produced during the verbal fluency tasks. FastText is a word embedding technique designed to capture the semantic meaning of words. It breaks words into character-level n-grams, which allows it to capture more contextual information and better handle rare or morphologically complex words. This makes FastText particularly effective for languages like Slovenian, as it can better represent linguistic nuances and provide more meaningful embeddings for semantic similarity analysis. Using these embeddings, we calculated cosine similarity, mean semantic distance, local optimal divergence, and global optimality divergence to capture the semantic relationships and coherence of word sequences.

By combining these analytical approaches, our study aims to provide a comprehensive understanding of the verbal fluency impairments associated with schizophrenia, contributing valuable insights to cognitive functioning.

4 RESULTS

The results of the verbal fluency tests conducted on both individuals with schizophrenia and healthy individuals are summarized in this chapter. Figure 1 and 2 display the most frequently spoken words by each group. Figure 1 presents the top five words spoken by healthy individuals, along with the occurrences of these words among people with schizophrenia. Similarly, Figure 2 illustrates the top five words spoken by individuals with schizophrenia, along with the occurrences of these words in the healthy group.

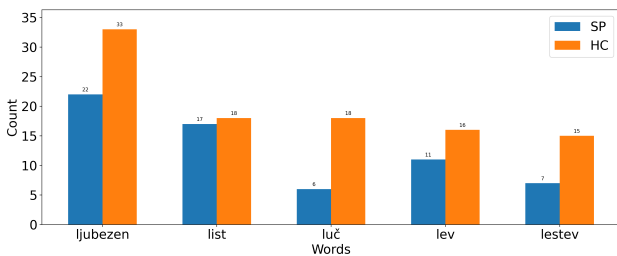


Figure 1: Top 5 Words by Healthy People.

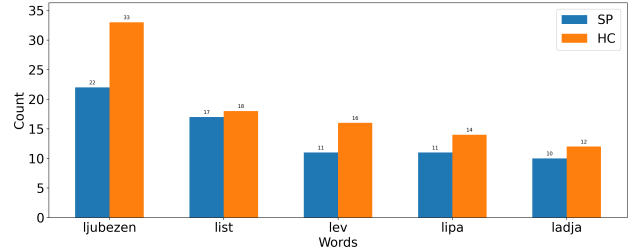


Figure 2: Top 5 Words by Individuals with Schizophrenia.

In addition to the graphical representation of the top words, Table 2 provides a summary of various key metrics from the verbal fluency tests. This includes the total number of different words, the number of unique words, and the counts of intrusion, repetition, and neologism words for both groups.

Table 2: Comparison of Verbal Fluency Test performance between individuals with schizophrenia (SP) and healthy controls (HC). The bottom rows summarize the total number of different words across all users and overlapping words between the groups.

| Total number of | SP | HC |
|---|-----|-----|
| Different words | 176 | 247 |
| Unique words | 60 | 131 |
| Intrusion words | 44 | 8 |
| Repetition words | 21 | 8 |
| Neologism words | 20 | 0 |
| Different words across all users | 307 | |
| Overlapping words between observed groups | 116 | |

The analysis indicates significant differences in verbal fluency between healthy individuals and those with schizophrenia. The following sections will provide a detailed examination of the data, including statistical analyses and further discussion on the implications of these findings.

4.1 Statistical Analysis

The statistical analysis compared the total number of words and the number of correct words (no intrusion, no repetition, no neologism) produced by participants in both groups.

Table 3: Mean and Standard Deviation (SD) of Words Produced by Healthy Controls (HC) and Individuals with Schizophrenia (SP).

| Measure | HC (Mean ± SD) | SP (Mean ± SD) |
|---------------|----------------|----------------|
| Correct Words | 10.32 ± 4.24 | 6.86 ± 4.19 |
| Total Words | 10.69 ± 4.32 | 8.52 ± 5.12 |

The results of the t-test, after confirming that the data is normally distributed, are summarized in the table below:

These results indicate significant differences between the groups, with healthy participants producing more correct and total words than participants with schizophrenia.

Table 4: T-test Results for Correct and Total Words by Group.

| Measure | t-statistic | p-value |
|---------------|-------------|---------|
| Correct Words | 4.77 | < 0.001 |
| Total Words | 3.04 | 0.002 |

4.2 Semantic Similarity Analysis

Using FastText embeddings, we calculated various semantic similarity measures. The average cosine similarity for the group with schizophrenia was 0.2, while for the healthy group, it was 0.19.

In addition to the average cosine similarity, we analyzed other measures which are summarized in Table 5.

Table 5: Semantic Similarity Measures.

| Measure | t-statistic | p-value |
|------------------------------|-------------|---------|
| Mean Semantic Distance | -0.59 | 0.554 |
| Global Optimality Divergence | 2.75 | 0.007 |
| Local Optimality Divergence | 0.29 | 0.769 |
| Repetitions | -2.26 | 0.026 |
| Intrusions | -1.83 | 0.070 |
| Neologisms | -4.47 | <0.001 |

The differences in global optimality divergence and occurrences of intrusions, repetitions, and neologisms were significant, highlighting disruptions in the overall semantic coherence and increased errors in individuals with schizophrenia.

5 DISCUSSION

The results of this study highlight significant differences in verbal fluency performance between individuals with schizophrenia and healthy controls. Healthy participants produced a higher number of correct words and exhibited more coherent and interconnected semantic structures compared to individuals with schizophrenia. These findings are consistent with established research on the cognitive impairments linked to schizophrenia, particularly in the domain of verbal fluency.

In addition to the differences in correct word production, individuals with schizophrenia also showed a significantly higher frequency of errors, including repetitions, neologisms, and intrusions. These errors are characteristic of the cognitive disorganization associated with schizophrenia and reflect the impaired executive function and semantic processing commonly observed in the disorder. The increased number of neologisms and intrusions further highlights the semantic and linguistic disruptions that differentiate individuals with schizophrenia from healthy controls.

Furthermore, although the use of FastText embeddings was a key aspect of this study, the method was not sensitive enough to capture local semantic disruptions. Measures such as cosine similarity and mean semantic distance, focusing on local semantic relationships, failed to highlight significant differences between the groups. This outcome suggests that while local word relationships may remain relatively preserved in individuals with schizophrenia, or that FastText embeddings may not effectively capture subtle local disruptions, the broader semantic coherence was impacted. This was evidenced by the significant differences in global optimality divergence, demonstrating a marked reduction in overall word sequence coherence among individuals with schizophrenia.

5.1 Conclusions

In conclusion, our study highlights the significant cognitive and linguistic impairments in individuals with schizophrenia, particularly in verbal fluency performance. We matched participants on age and gender to ensure comparability between the groups. While education is known to influence cognitive abilities, our analysis confirmed that within each educational level, there were no significant differences between the two groups. However, a larger sample size is needed to increase the power of our statistical analyses and allow for better generalizability and more in-depth exploration of all demographic variables.

Additionally, while FastText embeddings provided useful insights into semantic coherence, they were not sensitive enough to capture more subtle cognitive impairments. Future studies should explore alternative methods to provide a more comprehensive understanding of verbal fluency deficits in schizophrenia, contributing to improved diagnostic and therapeutic strategies.

REFERENCES

- [1] Flavia Galaverna, Adrián M. Bueno, Carlos A. Morra, María Roca, and Teresa Torralva. 2016. Analysis of errors in verbal fluency tasks in patients with chronic schizophrenia. *The European Journal of Psychiatry* 30 (12 2016), 305 – 320.
- [2] Kyrsten Grimes, George Fousias, Gary Remington, Kathryn Kalahani-Bargis, and Konstantine Zakzanis. 2020. Stability of Verbal Fluency in Outpatients with Schizophrenia. *Psychiatry Research* 295 (10 2020), 113528. <https://doi.org/10.1016/j.psychres.2020.113528>
- [3] Mary Kosmidis, Vassilis Bozikas, Christina Vlahou, Grigoris Kiosseoglou, George Giaglis, and Athanasios Karavatos. 2005. Verbal fluency in institutionalized patients with schizophrenia: Age-related performance decline. *Psychiatry research* 134 (05 2005), 233–40. <https://doi.org/10.1016/j.psychres.2005.02.003>
- [4] Nana Lehtinen, Ida Luotonen, and Anna Kautto. 2021. Systematic administration and analysis of verbal fluency tasks: Preliminary evidence for reliable exploration of processes underlying task performance. *Applied Neuropsychology: Adult* 30 (09 2021), 1–13. <https://doi.org/10.1080/23279095.2021.1973471>
- [5] Ojeda Natalia, Pedro Sanchez, Javier Peña, Edorta Elizagárate, Ana Yoller, Juan Larumbe, Miguel Gutiérrez-Fraile, Leonardo Casais, and Jesús Ezcurra. 2010. Verbal Fluency in Schizophrenia Does Cognitive Performance Reflect the Same Underlying Mechanisms in Patients and Healthy Controls? *The Journal of nervous and mental disease* 198 (04 2010), 286–91. <https://doi.org/10.1097/NMD.0b013e3181d61748>
- [6] Angel Nevado, David Del Río, María Teresa Martín-Aragoneses, José M. Prados, and Ramón López-Higes. 2021. Preserved semantic categorical organization in mild cognitive impairment: A network analysis of verbal fluency. *Neuropsychologia* 157 (2021), 107875. <https://doi.org/10.1016/j.neuropsychologia.2021.107875>
- [7] Dália Nogueira, Elizabeth Reis, and Ana Vieira. 2016. Verbal Fluency Tasks: Effects of Age, Gender, and Education. *Folia Phoniatrica et Logopaedica* 68 (12 2016), 124–133. <https://doi.org/10.1159/000450640>
- [8] Matthew M. Nour, Daniel C. McNamee, Yunzhe Liu, and Raymond J. Dolan. 2023. Trajectories through semantic spaces in schizophrenia and the relationship to ripple bursts. *Proceedings of the National Academy of Sciences* 120, 42 (2023), e2305290120. <https://doi.org/10.1073/pnas.2305290120>