The topic prominence in the speech of individuals with and without schizophrenia: an exploratory study

Ísis Fiorilo, Bruno Rocha

https://doi.org/10.1590/SciELOPreprints.15749

Submitted on: 2026-04-06
Posted on: 2026-05-04 (version 1)
(YYYY-MM-DD)

THE TOPIC PROMINENCE IN THE SPEECH OF INDIVIDUALS WITH AND WITHOUT SCHIZOPHRENIA: AN EXPLORATORY STUDY

Ísis Fiorilo
Mestranda em Linguística; Universidade Federal de Minas Gerais (UFMG); Belo Horizonte -MG; isisfiorilo@hmail.com; https://orcid.org/0009-0003-3203-9655.

Bruno Rocha
Doutor em Linguística; Universidade Federal de Minas Gerais (UFMG); Belo Horizonte -MG; bbruno791@gmail.com; https://orcid.org/0000-0001-7999-1545.

ABSTRACT

This is an exploratory study that investigates the prosodic realization of syllables with functional prominence in individuals with and without schizophrenia, focusing on f0 variation and its role in signaling functional prominence. Using the Language into Act Theory (L-AcT), the study analyzes the prosodic prominence of the informational unit of Topic, which establishes the cognitive domain for interpreting the illocution and is characterized by functional prosodic prominence. Results indicate that individuals with schizophrenia exhibit significantly lower f0 variation in nuclear syllables of topic units compared to those without schizophrenia. Specifically, speakers in the C-ORAL-ESQ corpus produced fewer instances of nuclear syllables with a four-semitone variation than those in the C-ORAL-BRASIL corpus.

Keywords: Schizophrenia, Language into Act Theory, Corpus Linguistics, Prosody.

INTRODUCTION

This study aims to investigate the prosodic realization of syllables with functional prominence in individuals with and without schizophrenia, to test the hypothesis that individuals with schizophrenia produce less f0 variation in syllables with functional prominence. Specifically, the study focuses on the analysis of the prominence of the informational unit of topic, as defined by the Language into Act Theory (Cresti, 2000; Moneglia and Raso, 2014; Cavalcante, 2020). This is an exploratory study in which occurrences from a limited number of speakers from each group (with and without schizophrenia) will be analyzed. The goal is to establish the basis for a more in-depth analysis in the future, which will rely on a larger number of speakers.

The Language into Act Theory is an approach developed to analyze the informational structure of spontaneous speech, based on the idea that each linguistic sequence that is perceived as prosodically concluded features one or more informational patterns, each of which performs a speech act directed at the interlocutor. An informational pattern consists of one or more tonal units, where one serves an illocutionary function, while the others may

fulfill various subordinate functions to the illocutionary unit. The topic, one of the non-illocutionary units that may compose an informational pattern, establishes a cognitive domain for the interpretation of the illocution and is marked by a functional prosodic prominence.

This research is inspired by a study by Martínez-Sánchez *et al*. (2015), which found significant differences in local f0 measures when comparing the speech of individuals with and without schizophrenia (for example, the number of syllables with variations equal to or greater than 4 semitones), although no significant differences were observed in global measures (such as the mean and standard deviation of f0 throughout the utterance). Additionally, Saccone's (2023) study employs measures similar to those we aim to implement and identified significant differences between individuals with and without schizophrenia in Italian.

Based on these findings, we formulate the following hypotheses: 1. Individuals with schizophrenia produce fewer syllables in the topic unit with a difference equal to or greater than 4 ST than individuals without schizophrenia; 2. Individuals with schizophrenia produce fewer nuclear syllables in the topic unit with a difference equal to or greater than 4 ST than individuals without schizophrenia.

LANGUAGE INTO ACT THEORY

The Language into Act Theory (Cresti, 2000; Moneglia and Raso, 2014; Cavalcante, 2020), or L-AcT, is a pragmatic approach developed to analyze the informational structure of spontaneous speech, positioning itself as an heir to Austin's (1962) Speech Act Theory. According to this theory, every prosodically completed linguistic sequence corresponds to a speech act formed by one or more informational patterns, each performing an illocutionary act directed at the interlocutor.

The informational pattern necessarily comprises an illocutionary unit and may contain non-illocutionary units with various functions, all conveyed prosodically and serving as support for the illocution. Among the non-illocutionary units, the topic is the most frequent and the most important. It functions to establish a cognitive domain for the illocution and is necessarily located before the illocutionary unit. Prosodically, the topic has a functional nucleus, i.e., a set of syllables that exhibit specific prosodic properties responsible for signaling the function of the unit. The functional nucleus of the topic can be realized in three different ways:

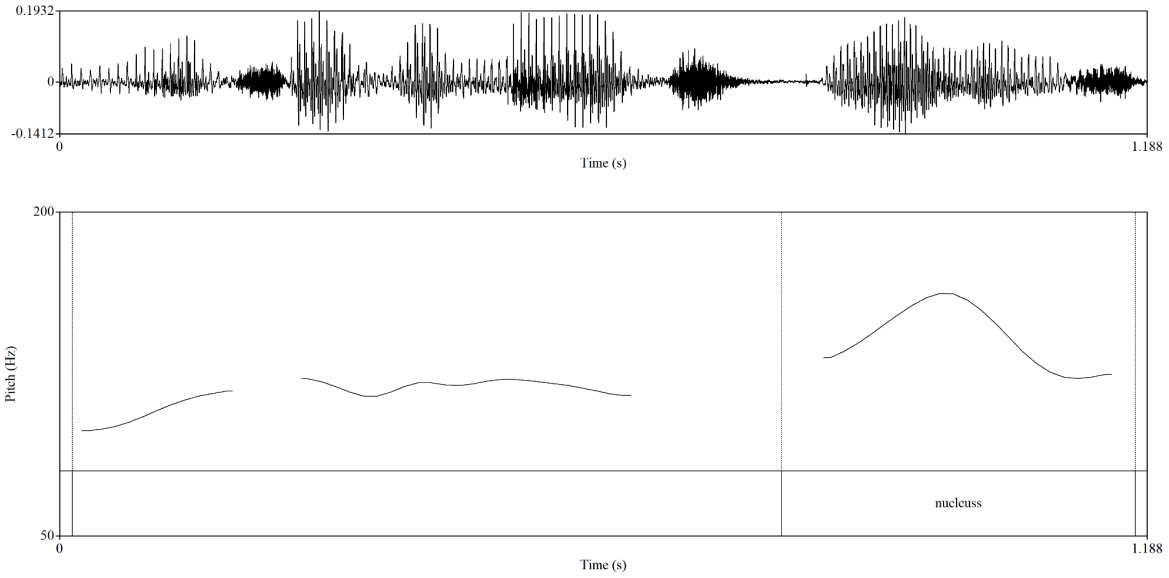
- Type 1: rise-fall f0 contour aligned with the last stressed syllable of the unit;

- Type 2: rising f0 contour aligned with the last stressed syllable;
- Type 3: two semi-nuclei; the first presents high/extra high f0 values, and the second features lower f0 values combined with syllable lengthening.

In cases where the informational unit of topic contains more syllables than necessary to realize the functional nucleus, the additional syllables do not carry informational functions and are referred to as preparation. In Types 1 and 2, preparation is always located at the beginning of the unit, before the nuclear syllables. In Type 3, there may also be preparation syllables between the two semi-nuclei.

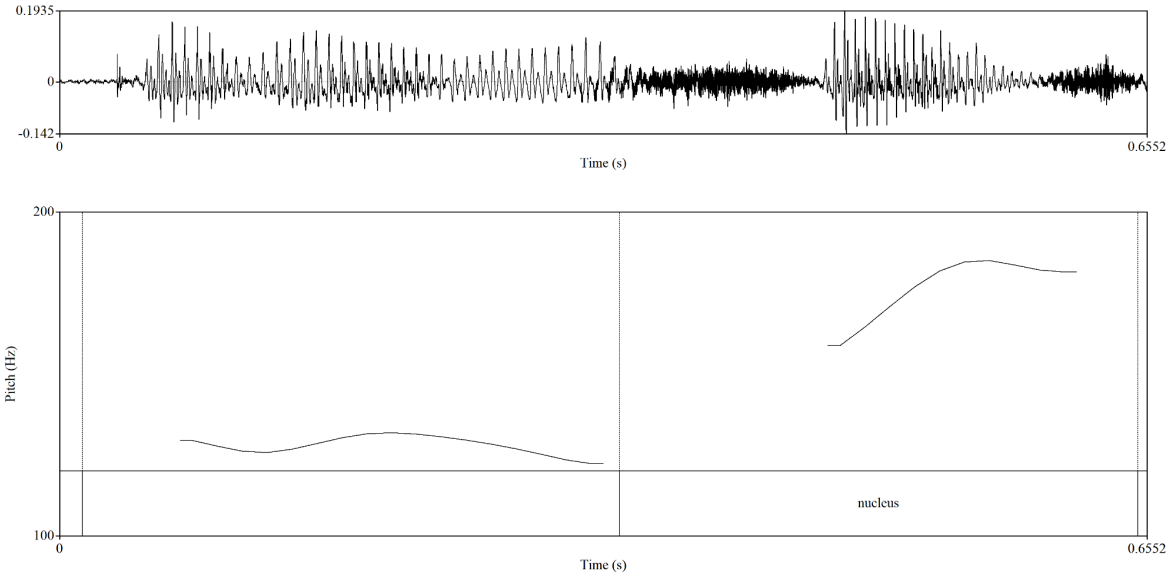
The following examples, extracted from the American English minicorpus (Cavalcante, 2016), begin with a topic unit, which precedes illocutionary units. The figures illustrate the f0 contour of the topic units in each example, indicating the nucleus of each.

Example 1 (afamdl01_80): once I get my ex**perience** /=TOP= I'll be up there too / in the top-four salesmen //



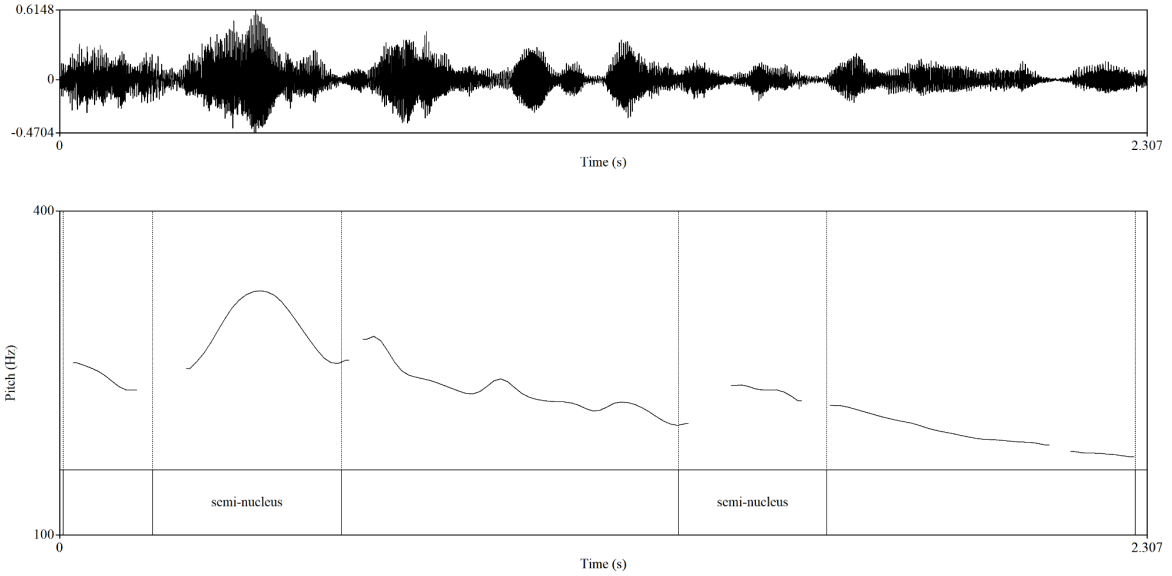
Subtitle: contour of the topic unit in Example 1. The nucleus corresponds to the syllables "perience" of the word "experience", in bold in the example. The previous syllables are the preparation for the nucleus.

Example 2 (afamdl01_67): but in a **sense** /=TOP= I need a / some type of steady income //



Subtitle: F0 contour of the topic unit in Example 2. The nucleus corresponds to the word “sense”, in bold in the example. The previous syllables are the preparation for the nucleus.

Example 3 (afamcv04_138): when **Mary** tells me to get **a sleep** over the weekend /=TOP= you know I need to get sleep over the weekend //



Subtitle: F0 contour of the topic unit in Example 2. The first semi-nucleus corresponds to the word “Mary”, and the second semi-nucleus corresponds to the NP “a sleep”, in bold in the example. All the other syllables are the preparation for the semi-nuclei.

METHODS

Corpora

Two corpora were used for this work, the C-ORAL-BRASIL (Raso & Mello, 2012) and the C-ORAL-ESQ (Author, 2025). The C-ORAL-BRASIL corpus is a reference corpus for spoken Brazilian Portuguese. The C-ORAL-ESQ corpus, documents the speech of people

with schizophrenia during psychiatric consultations. Furthermore, the two corpora are prosodically segmented in terminal and non-terminal boundaries, and have smaller sections that are informationally annotated according to L-AcT.

The C-ORAL-BRASIL corpus was compiled following the same criteria established for C-ORAL-ROM corpora (Cresti and Moneglia, 2005), and comprises 139 texts distributed in two different contexts: public and private. For each context, there is a subdivision of monological, dialogical, and conversational texts, resulting in a total of 208,130 words (Raso and Mello, 2012). To guarantee great diaphasic variation, C-ORAL-BRASIL I has texts recorded in diverse situations, representing the open universe of spontaneous speech in the most reliable way possible. Besides that, the corpus was morphosyntactically labeled using the PALAVRAS parser (Bick *et al*., 2000), in addition to text-sound alignment in the WinPitch software (Martin, 2004).

The C-ORAL-ESQ corpus (Author, 2025) consists in 43 recordings of psychiatric consultations between patients with schizophrenia and their respective physicians at two public institutions in Belo Horizonte, Minas Gerais: the Raul Soares Institute (IRS-FHEMIG) and the Hospital das Clínicas of the Federal University of Minas Gerais (HC/UFMG). Of the consultations included in C-ORAL-ESQ, 19 were recorded in both audio and video, while the remaining 24 were recorded in audio only. The corpus contains a total of 103,633 words, including 47,808 words produced by patients, 48,069 by physicians, 6,911 by companions, and 842 by other participants.

It is worth noting that all recorded consultations are part of ongoing psychiatric treatment and were conducted by the same physicians who usually carry them out, in the same consulting rooms where they normally take place, thus ensuring ecological validity for the data. C-ORAL-ESQ also includes metadata with general information about all participants, as well as detailed sociodemographic and clinical information for the 43 patients. This includes the results of psychometric scales such as the PANSS (Kay, S. R.; Fiszbein, A.; Opler, L. A., 1987), which assesses the positive and negative symptoms as well as the general psychopathology of schizophrenia, and the BACS (Keefe *et al*., 2006), which measures the cognitive impairments associated with the disorder. C-ORAL-ESQ underwent the same data processing procedures applied to C-ORAL-BRASIL and includes text–sound alignment (and, when available, text–sound–video alignment) in the ELAN software (Max Planck Institute for Psycholinguistics, n.d.), as well as morphosyntactic tagging using the Palavras parser (Bick *et al*., 2000).

Additionally, the C-ORAL-MED, a multimodal comparison corpus for C-ORAL-ESQ, is currently being compiled. It will document medical consultations involving individuals without mental disorders or speech pathologies. The C-ORAL-MED will be fully multimodal and will include at least the same number of recordings as C-ORAL-ESQ. Furthermore, the corpus will be matched, as far as possible, with C-ORAL-ESQ in terms of participants, sex, age, and level of education.

Participants Overview

The table below presents demographic information and PANSS clinical scores for each subject, including both the positive (PANSS-P) and negative (PANSS-N) syndrome scales.

Table 1: Demographics and PANSS Clinical Scores.

Subject	Sex	Age	PANSS-P	PANSS-N
ESQ1	M	C	9	11
ESQ2	F	C	11	17
ESQ3	F	C	15	35
ESQ4	M	B	19	10
ESQ5	M	C	17	19
ESQ6	M	B	8	22
BR1	M	C	n/a	n/a
BR2	F	D	n/a	n/a
BR3	M	D	n/a	n/a
BR4	F	B	n/a	n/a
BR5	M	C	n/a	n/a
BR6	M	C	n/a	n/a

Source: Authors.

The *subject* refers to the identifier for each participant in the study. Sex denotes the participant's gender, where M stands for Male and F for Female. The *age* category is classified into three groups: B representing individuals aged 26-40 years, C for those aged 41-60 years, and D for individuals aged 60 years and older. The PANSS scale (Positive and Negative Syndrome Scale) is employed to assess the severity of symptoms in individuals

with schizophrenia. It evaluates both positive symptoms, such as hallucinations and delusions, and negative symptoms, including blunted affect and lack of motivation.

The sample consists of 8 male participants and 4 female participants. The *age* distribution is as follows: 3 participants are classified in age group B (26-40 years), 7 participants fall within age group C (41-60 years), and 2 participants belong to age group D (60 years and older). The *PANSS-P* (positive symptoms) and *PANSS-N* (negative symptoms) scores exhibit variability across participants. For instance, subject ESQ1 presents relatively mild symptom severity, with a *PANSS-P* score of 9 and a *PANSS-N* score of 11, while subject ESQ3 demonstrates more pronounced symptomatology, with a *PANSS-P* score of 15 and a *PANSS-N* score of 35. This variability underscores the heterogeneous nature of symptom severity within the sample. Considering that all patients included in the C-ORAL-ESQ corpus are currently undergoing treatment and are clinically stabilized, none presents a PANSS-P score higher than 19, which was the cutoff criterion adopted for the corpus.

Topic Units and Acoustic Measurements

As a preliminary study aimed at establishing the methodology for a more comprehensive future analysis, a limited sample of speakers and topic units was analyzed.We extracted 44 topic units from 6 speakers from C-ORAL-ESQ and 47 topic units from 6 speakers from C-ORAL-BRASIL. The topic units from C-ORAL-ESQ had 281 syllables in total (123 nuclear, and 158 non-nuclear), while the topics from C-ORAL-BRASIL had 261 syllables (92 nuclear and 169 non-nuclear).

For each syllable, we extracted minimum and maximum f0 values (in semitones-ST) and duration (in milliseconds) and calculated the f0 variation (f0 max - f0 min) and the f0 variation rate (f0 variation/duration). The following measurements were also extracted: number of syllables with f0 variation equal or greater than 4 ST; mean, median, minimum, and maximum f0 variation; mean, median, minimum, and maximum f0 variation rate (f0 variation/duration). All measurements were manually exctracted using Praat software (Boersma and Weenink, 2025).

RESULTS

The following tables show the measurements obtained for nuclear and non-nuclear topic syllables in C-ORAL-ESQ and C-ORAL-BRASIL.

Table 2: Nuclear syllables statistics for C-ORAL-ESQ and C-ORAL-BRASIL corpora.

Corpus	**Count**	**f0 variation**		**f0 var. rate**

		Mean	SD	>=4ST	Mean	SD
ESQ	123	3.1	2.1	23 (19%)	18.6	15.8
BR	92	3.5	2.5	38(41%)	19.9	18.0

Source: Authors.

Subtitle: Description of the acoustic measures used in the analysis. f0 variation corresponds to the difference between minimum and maximum f0 values (f0 min – f0 max). ≥4 indicates the number of syllables with f0 variation equal to or greater than 4 semitones (ST). f0 var. rate corresponds to the ratio between f0 variation and duration (f0 variation / time).

As seen in Table 2, the C-ORAL-ESQ corpus presents 23 nuclear syllables with a variation of f0 ≥ of 4 semitones (19% of 187 nuclear syllables), while the C-ORAL-BRASIL corpus presents 38 nuclear syllables with the same variation of f0 (41% of the nuclear syllables). This difference was statistically significant (Chi-square test: χ 2 (1) = 12.15, p = 0.00049).

Table 3 shows the f0 variation and f0 variation rate measurements for the non-nuclear syllables. As expected, both corpora present fewer non-nuclear syllables with a variation of f0 ≥ 4 semitones compared to nuclear syllables, since non-nuclear syllables do not have functional prominences. A variation of f0 ≥ 4 semitones can be observed in only non-nuclear 14 syllables of C-ORAL-ESQ (9% of the total) and in 30 syllables of C-ORAL-BRASIL (18%). Consistently, the results were statistically significant (Chi-square test: χ 2 (1) = 4.81, p = 0.028).

Table 3: Non-nuclear syllables statistics for C-ORAL-ESQ and C-ORAL-BRASIL corpora.

Corpus	Count	f0 variation			f0 var. rate	
		Mean	SD	>=4ST	Mean	SD
ESQ	158	2.2	1.8	14 (9%)	16.6	21.2
BR	169	2.3	1.4	30 (18%)	15.9	12.1

Source: Authors.

Subtitle: Description of the acoustic measures used in the analysis. f0 variation corresponds to the difference between minimum and maximum f0 values (f0 min – f0 max). ≥4 indicates the number of syllables with f0 variation equal to or greater than 4 semitones (ST). f0 var. rate corresponds to the ratio between f0 variation and duration (f0 variation / time).

Furthermore, Wilcoxon rank-sum tests revealed significant differences between C-ORAL-ESQ and C-ORAL-BRASIL in all measures analyzed: nuclear syllables' f0 variation (W = 4310, p = 0.0028), nuclear syllables' f0 variation rate (W = 4197, p = 0.0012), non-nuclear syllables' f0 variation (W = 0.292, p = 0.00034), and non-nuclear syllables' f0

variation rate (W = 9168, p = 0.000001). These results suggest a significant difference in f0 variation patterns between the two corpora.

Although the results of the Chi-square and Wilcoxon rank-sum tests are statistically significant, they should be interpreted with caution. Given the exploratory nature of this study, all statistical tests were conducted with a very limited number of speakers. Therefore, these results should be regarded primarily as an indication of potential differences in the observed values between the two corpora, rather than as definitive evidence of broader patterns. Future analyses based on a larger sample of speakers will be necessary to confirm the tendencies identified in this preliminary investigation.

DISCUSSION

This study investigated the prosodic realization of topic units in the speech of individuals with and without schizophrenia, focusing on f0 variation and its role in signaling functional prominence. Our findings indicate that individuals with schizophrenia exhibit significantly lower f0 variation in nuclear syllables of topic units compared to individuals without schizophrenia. Specifically, speakers in the C-ORAL-ESQ corpus produced fewer nuclear syllables with f0 variations equal to or greater than 4 semitones than those in the C-ORAL-BRASIL corpus. This pattern was also observed in non-nuclear syllables, where a significant difference in f0 variation was found between the two corpora. However, it should be emphasized that this was an exploratory study based on a small number of speakers; therefore, the statistical significance of these results should be interpreted with caution.

Future research should expand the dataset to include a larger sample of speakers and explore additional prosodic features, such as speech rhythm and intensity variation. Future studies will also compare the C-ORAL-ESQ corpus with the C-ORAL-MED corpus, allowing for the analysis of speech produced in an analogous communicative context, namely medical consultations. Investigating the relationship between prosodic markers and clinical variables, such as symptom severity and medication effects, could further enhance our understanding of how schizophrenia influences spoken language.

RESEARCH DATA AVAILABILITY

The datasets analyzed during the current study are available on the C-ORAL-BRASIL website (https://www.c-oral-brasil.org/).

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AI USE DISCLOSURE

The authors declare that no artificial intelligence tools were used in the writing or editing of this manuscript.

AUTHORS' CONTRIBUTION

All authors made substantial contributions to the conception and design of the study, data analysis and interpretation, and the revision of the manuscript and approved the final version. All authors assume public responsibility for the content of the manuscript.

REFERENCES

Author, 2025.

AUSTIN, J. L. How to do things with words. Oxford: Oxford University Press, 1962.

BICK, E. et al. The annotation of the C-ORAL-BRASIL oral corpus through the implementation of the Palavras parser. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION (LREC), 8., 2012. Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12). Istanbul: ELRA, 2012. p. 3382–3386.

BOERSMA, P.; WEENINK, D. Praat: doing phonetics by computer. Version 6.4.27. 2025. Disponível em: http://www.praat.org/. Acesso em: 27 jan. 2025.

CAVALCANTE, F. A. The informational unit of topic: a crosslinguistic, statistical study based on spontaneous speech corpora. 2020. Tese (Doutorado em Estudos Linguísticos) – Universidade Federal de Minas Gerais, Belo Horizonte, 2020.

CAVALCANTE, F.; RAMOS, A. The American English spontaneous speech minicorpus. CHIMERA: Romance Corpora and Linguistic Studies, v. 3, n. 2, p. 99–124, 2016.

CRESTI, E. Corpus di italiano parlato. v. 1–2. Firenze: Accademia della Crusca, 2000. CD-ROM.

CRESTI, E.; MONEGLIA, M. (eds.). C-ORAL-ROM: integrated reference corpora for spoken Romance languages. Amsterdam: John Benjamins, 2005.

ELAN. ELAN (Version 6.9) [computer software]. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive, 2024. Disponível em: https://archive.mpi.nl/tla/elan. Acesso em: 27 jan. 2025.

KAY, S. R.; FISZBEIN, A.; OPLER, L. A. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophrenia Bulletin, v. 13, n. 2, p. 261–276, 1987.

KEEFE, R. S.; POE, M.; WALKER, T. M.; KANG, J. W.; HARVEY, P. D. The Schizophrenia Cognition Rating Scale: an interview-based assessment and its relationship to cognition, real-world functioning, and functional capacity. American Journal of Psychiatry, v. 163, n. 3, p. 426–432, 2006.

MARTIN, Philippe. WinPitch: a tool for speech analysis and intonation modelling. In: SPEECH PROSODY, 2004, Nara. Proceedings of Speech Prosody 2004. Nara: International Speech Communication Association, 2004.

MARTÍNEZ-SÁNCHEZ, F. et al. Can the acoustic analysis of expressive prosody discriminate schizophrenia? The Spanish Journal of Psychology, v. 18, p. 1–9, 2015.

MAX PLANCK INSTITUTE FOR PSYCHOLINGUISTICS. ELAN (Version 6.9) [Computer software]. Nijmegen: The Language Archive, 2023. Disponível em: https://archive.mpi.nl/tla/elan. Acesso em: 10 mar. 2026.

MONEGLIA, M.; RASO, T. Notes on the Language into Act Theory. In: MONEGLIA, M.; RASO, T. (eds.). Spoken corpora and linguistic studies. Amsterdam; New York: John Benjamins, 2014. p. 468–489.

RASO, T.; MELLO, H. (eds.). C-ORAL-BRASIL I: corpus de referência do português brasileiro falado informal. Belo Horizonte: Editora UFMG, 2012.

SACCONE, V.; TRILLOCCO, S.; MONEGLIA, M. Markers of schizophrenia at the prosody/pragmatics interface: evidence from corpora of spontaneous speech interactions. Frontiers in Psychology, v. 14, p. 1233176, 2023. Disponível em: https://doi.org/10.3389/fpsyg.2023.1233176. Acesso em: 27 jan. 2025.

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.