



The C-ORAL-ESQ project: a corpus for the study of spontaneous speech of individuals with schizophrenia

Tommaso Raso¹ · Bruno Neves Rati de Melo Rocha¹ ·
João Vinícius Salgado¹ · Breno Fiuza Cruz¹ · Lucas Machado Mantovani² ·
Heliana Mello¹

Accepted: 29 May 2023 / Published online: 27 June 2023
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract

This paper presents the C-ORAL-ESQ corpus project, which is dedicated to the study of the speech of individuals with schizophrenia. The main aim of the project is to investigate cognitive aspects of individuals with schizophrenia. This investigation is carried through the compilation of a spontaneous speech corpus and its study, which focuses mainly on the analysis of information structuring and its prosodic correlates. The paper mainly deals with the methodological aspects of the corpus compilation and reports its present stage: it informs about the environment and the setting of the sound file recordings, the medical and ethical criteria for the selection of the participants, the corpus aimed dimensions and the present stage of compilation, as well as its design and compilation criteria, which include attention to prosodic annotation, and metadata related to the participants' characteristics. Additionally, the theory adopted for the study of information structure is summarized, focusing on those aspects that can better address cognitive processes of individuals with schizophrenia and their prosodic correlates. Finally, the perspectives for future studies and resource compilations are presented.

Keywords Schizophrenic speech · Corpus · Information structure · Prosody

1 Introduction

In this paper we present a Brazilian spoken corpus of individuals with schizophrenia, called C-ORAL-ESQ (Corpus Oral de Esquizofrênicos). We describe the environment and the setting in which data are collected, the medical and

✉ Tommaso Raso
tommaso.raso@gmail.com

¹ UFMG, Belo Horizonte, Brazil

² IRS and FCMMG, Belo Horizonte, Brazil

ethical criteria for participants' inclusion, the corpus compilation methodology, the current stage of data compilation and participants' characteristics. Additionally, we provide a brief summary of the studies that have already been produced within the project, along with some information about the theoretical framework adopted, and the perspectives that we have for further studies and the compilation of new data.

The main goal of the project is to study the cognition of individuals with schizophrenia through their linguistic performance, mainly investigating how they manage information structure and its prosodic correlates. The C-ORAL-ESQ project is carried at the Federal University of Minas Gerais (UFMG) and the *Instituto Raul Soares* (IRS), by a team of psychiatrists and linguists. The project is coordinated by Raso and Salgado.

The remainder of this paper is organized as follows: Sect. 2 describes the health facility where data is collected, the protocols followed in patients' treatment and the criteria for patients' inclusion in the corpus. Additionally, some information about the planning of a control corpus is also presented. Section 3 is devoted to the corpus description and its compilation methodology, including the transcription and prosodic annotation criteria, in addition to participants' metadata profiling. Section 4 explains the criteria adopted for the informational annotation, introducing some basic aspects of the theory that guides the tagging process, in addition to summarizing the first results of studies carried within C-ORAL-ESQ. Finally, the perspectives for future work are presented.

The use of clinical databases is crucial for the study as well for early diagnosis and treatment of mental disorders. Text corpora limit the nature of studies that can be carried, since they cannot account for the enormous amount of information conveyed by the sound signal, including prosody. It is mainly prosody, in fact, that conveys emotion, attitude, illocution and information structure. Nevertheless, spoken corpora, and especially spontaneous speech ones, are still rare, and absent in several important languages (see Li et al., 2019, for a survey; see also Forjó et al., 2022, for a resource in European Portuguese), while it is well known that several speech measures can function as a biomarker in mental diseases (Rapcan et al., 2010).

There is strong evidence that schizophrenia affects both the expression and the perception of prosody. Several studies show that the production and recognition of emotional prosody is impaired with strong statistical relevance, as shown by a meta-review by Hoekert and Aleman (2007). There is also evidence that the disease affects prosodic parameters such as intensity (Cohen et al., 2013; Compton et al., 2018), pause (Cohen et al., 2013; Martínez-Sánchez et al., 2015), and f_0 (Alpert et al., 2000; Cohen et al., 2013; Compton et al., 2018). Regarding the discourse level, schizophrenia is associated with alterations on the information structure (Cresti et al., 2005) and the production of contrastive focus (Michelas et al., 2014). Finally, corroborating the need for a better understanding of the prosodic perception and expression of individuals with schizophrenia, various psychometric scales, such as PANSS (Kay et al., 1987) and SCoRS (Keefe et al., 2006), point (in some cases explicitly) to alterations in the speech of the patients.

2 Health facilities and criteria for the inclusion of subjects in the corpus

Individuals enrolled in this study have the diagnosis of schizophrenia and, at the time of their recording, all of them were undergoing treatment at Instituto Raul Soares. This facility is a centenary public state hospital in Belo Horizonte (Brazil), specialized in mental health. Additionally, it is an educational centre, provided with a long-lasting full time medical residency in psychiatry.

Usually, patients reach to the hospital for the first time due to a psychotic crisis and take brief periods of treatment in the inpatient wards. When their health improves, they are referred to a monthly or bimonthly assessment in the outpatient clinic by a resident doctor (under supervision). The outpatient clinic of Instituto Raul Soares has a division of competence, and patients with schizophrenia have their own setting coordinated by a senior psychiatrist and researcher. As expected in a medical educational setting, resident doctors change on a regular basis (often yearly), but the treatment rational stays the same as warranted by the supervisor.

Antipsychotic medication is a cornerstone in the treatment of patients with schizophrenia, both in the period of crisis as in the maintenance phase (stability) (Lieberman, 2005). Nonetheless, they are effective only over psychotic symptoms (positive symptoms), leaving unattended the remaining core symptoms – negative (Galderisi, 2018) and cognitive (Kahn & Keefe, 2013), particularly.

Patients in the outpatient clinic are typically in a phase of illness stability, described mainly by attenuated psychotic symptoms (low positive symptoms) and by a variable degree of negative symptoms—poor affective expression and motivational drive. This clinical profile is the most important inclusion criterion of this study—patients with low positive symptoms (stability), under antipsychotic treatment. Patients invited to join the research need to have its aims and methods clear, and when in agreement with its terms, are required to sign a written informed consent in accordance with the requirements of the Ethics Committee for Research and the Declaration of Helsinki (World Medical Association, 2013).

The diagnosis of schizophrenia is clinically established, and for research purposes, it is confirmed through the Structured Clinical Interview for DSM-5 (SCID-5) (First, 2016). Numerically, stability is determined by using the Positive and Negative Symptoms Scale (PANSS) (Kay, 1987)—patients scoring 19 or less in the Positive Subscale, and 4 or less in any item of the same subscale. PANSS is also useful in the assessment of positive and negative symptoms. Sociodemographic data of patients is assessed using a semi-structured instrument developed and tested by our group, which includes age, educational level, occupation, marital status, gender, birthplace and socioeconomic status. Similarly, clinical data is obtained through a semi-structured instrument covering age of disease onset, duration of illness, psychiatric medication in use (class, dose, purpose), age of treatment onset, first psychotic episode (yes/no), number of hospital admissions, history of drug (or tobacco or alcohol) use, clinical diseases, non-psychiatric medication in use, family psychiatric history. Measurement of neurological and

motor side effects due to use of antipsychotic medication is part of the research protocol, and the following instruments are employed: Simpson-Angus Scale (parkinsonism) (Simpson & Angus, 1970, Louzã Neto, 1998), Abnormal Involuntary Movements Scale (AIMS; tardive dyskinesia) (Louzã Neto, 1998) and Barnes Akathisia Rating (Barnes, 1989, Louzã Neto, 1998).

Alongside the basic clinical information listed in the previous paragraph, the research protocol also includes a more complex measurement of cognitive performance (a short version of the Brief Assessment of Cognition in Schizophrenia–BACS-SF) (Lam et al, 2017), social cognition (Hinting Task and Bell-Lysaker Emotion Recognition Task) (Bryson et al, 1997; Corcoran et al, 1995), functional capacity (a brief version of the UCSD Performance-based Skills Assessment–UPSA-B) (Mausbach, 2007) and real-world functioning (Personal and Social Performance–PSP) (Morosini, 2000).

Both basic and complex assessments have the purpose of describing in the most reliable way the clinical profile of patients enrolled in the research. Furthermore, this data serves as an important parameter of analysis of future linguistic and prosodic information, when groups of patients can be studied separately (concerning some clinical outline), or when groups of symptoms can be controlled as covariates. For example, we hypothesize that in a certain way, deficits in facial expression (a core negative symptom–Affect) or in the ability of decoding facial expression (a core domain of social cognition) may be related to some sort of prosody profile.

In the upcoming months, the subjects of this study will be extended to the Hospital das Clínicas of the Federal University of Minas Gerais (HC-UFGM). HC-UFGM is a teaching hospital linked to a Brazilian public university. The hospital offers outpatient care to people with severe mental disorders from across the state of Minas Gerais. Its psychiatry service receives requests for psychiatric evaluation of outpatient care from other medical specialties of the hospital, as well as external requests. The services are linked to the Brazilian Unified Health System. The schizophrenia outpatient clinic provides psychiatric and nursing care to people diagnosed with schizophrenia and related disorders. The third-year residents of the hospital's psychiatric residency program perform medical care, supervised by a psychiatry professor from the Mental Health Department at the UFGM School of Medicine. Patients treated at the outpatient clinic also have guidelines on how to maintain good health practices, such as sleep hygiene, healthy eating and physical activity. These orientations are carried out by undergraduate students supervised by a senior professional. The clinical characteristics of the patients treated are similar to those found at Instituto Raul Soares.

Additionally, in the upcoming months, subjects will be recruited for a control group to be compared with the patients with schizophrenia. These subjects will be patients from both general medicine and gastroenterology outpatient services from HC-UFGM. The medical care in these services is provided by undergraduate medical school students (third year of the medical course or more) under supervision of a senior medical school professor. It should be noted that a typical psychiatric interview is longer and contains much more verbal interplay compared to a typical general medical interview conducted by a senior physician. However, when conducted by an ungraduated student, a general medical interview becomes similar to

Table 1 Aimed at and current word statistics of the C-ORAL-ESQ corpus

	Recordings	Patients	Words by patients (mean)	Words by patients (total)	Total words	Length (hs)
Aimed	40	40	1500	60,000	–	–
Current	30	28	1505 (SD 1,014)	45,161	90,903	09:21:32

a psychiatric interview in terms of speech speed and amount of verbal interplay. Therefore, we believe that the conditions in a general medical consultation with an undergraduate student will be quite similar to those from a psychiatric consultation with a psychiatry resident. Exclusion criteria for control subjects include any major psychiatric or neurological disorder, as well as any linguistic pathology. The control group will be matched to the schizophrenia group according to age, sex, and education.

3 The C-ORAL-ESQ corpus

3.1 Corpus dimension and basic features of the recordings

The C-ORAL-ESQ corpus aims to study the spontaneous speech of individuals with schizophrenia by documenting regular psychiatric consultations between patients and their physicians. Upon completion, the corpus will comprise at least 40 recordings with approximately 1,500 words each, uttered by the patients (plus the words uttered by the physicians, the patients' accompanying persons and intervenient interactions that might have taken place, which are also recorded and fully transcribed).

Currently, the corpus is comprised of 30 recordings from 28 different patients (2 of them have been recorded twice¹), with a mean of 1,505 words per patient, and a total of 45,161 words produced by patients and 90,903 words in total. The aimed at and the present word statistics are shown in Table 1.

Table 2 displays the current stage of the procedures that are employed in the C-ORAL-ESQ corpus compilation. Due to their length, nine recordings were divided into two different excerpts, so at the current time the corpus features 39 different excerpts.² Only three of them have not been transcribed yet. Two excerpts were transcribed and prosodically segmented (a procedure always carried out during

¹ Sometimes the logistics of the recordings allows for a second recording, at a distinct occasion, with previously recorded patients. We think this might be useful, since it allows for intrasubject analysis at different stages of the treatment. The inclusion of more than one recording of the same patient in the corpus will be decided at the end of the compilation process. It might be useful to have a special session dedicated to this specific type of recording. In any case, the corpus foresees the recording of 40 patients.

² This helps us to have excerpts of different interactional typologies, some more monological and some more dialogical. This difference impacts several speech measurements and informational characteristics of the corpus, and should therefore be taken into account whenever it is possible (see Cresti, 2005; Raso and Mittmann, 2012).

Table 2 Corpus compilation procedures

	Excerpts	Words by patients	Total words	% words
Not transcribed	3	–	–	–
Transcribed	2	2645	4012	0%
Transcription revised	11	8139	23,827	29%
Aligned	10	12,599	25,050	32%
Alignment revised	3	3826	7982	0%
Tagged	2	3362	4696	5%
Tagging revised	7	11,018	18,343	20%
TOTAL	39	45,161	90,903	–

Table 3 Number of words produced by patients, health professionals, patients' accompanying person and intervenient participants in the recordings

	Mean	SD	Median	Min	Max	Total
All participants	3030	1525	2741	1055	8711	90,903
Patients	1505	1014	1639	101	5193	45,161 (50%)
Health prof	1325	800	1082	554	3518	39,741 (44%)
Accompanying	198	287	0	0	851	5935 (6%)
Intervenient	2	10	0	0	56	66 (0%)
Length (hs)	00:15:36	00:07:54	00:13:26	00:04:58	00:38:56	09:21:32

the transcription, see Sects. 3.2 and 3.3). Eleven excerpts had their transcription and segmentation revised by expert annotators. Ten excerpts were also sound-to-text aligned through WinPitch (Martin, 2015) and/or ELAN (Version 6.4, 2022; Wittenburg et al., 2006)³ software, and another three have already had their alignment revised. Nine excerpts have been informationally tagged according to the Language into Act Theory (L-AcT) (Cresti, 2000; Moneglia; Raso, 2014) and seven of them have been revised during discussion sessions among the annotators.

Approximately 1/3 of the corpus that has been collected so far presents a revised transcription (29%), another 1/3 has also been aligned (32%), and the last 1/3 has already been tagged (5%) or tagged and revised (20%), thus being ready for all kinds of studies.

Table 3 specifies the number of words produced by each group of participants (patients, health professionals, accompanying persons, intervenient individuals), which allows the drawing of some important insights on the structure of the consultations. First of all, it is important to notice that, among the 90,903 words that compose the corpus, 96% of them were produced by patients and physicians, and

³ The alignment process initially used WinPitch. When we received permission to record multimodally, we decided to switch to ELAN. All the previous alignments with WinPitch are being converted to ELAN.

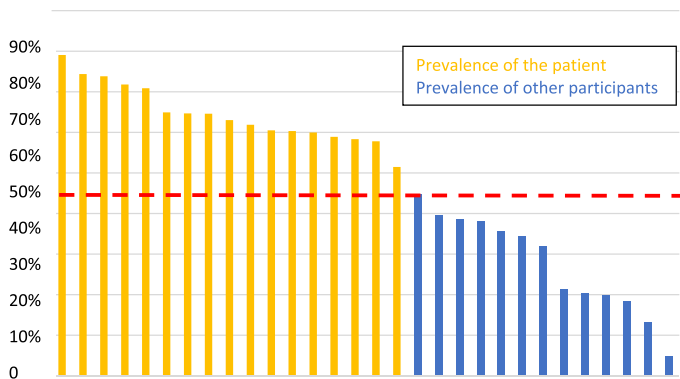


Fig. 1 Percentage of words produced by patients in relation to the words produced by other participants (health professionals, accompanying person and intervenient participants)

only 6% by the patients' accompanying persons. Intervenient participants produced a total of 66 words, which correspond to 0.07% of the total.

Based only on a superficial look at these data, it could be inferred that the participants of the consultations act homogeneously among them. However, the high SD in the number of words produced by patients, health professionals and also accompanying individuals, as well as the median of these word counts, suggest a different picture: while in most consultations there is a balance between the number of words uttered by patients and physicians, there are also recordings in which the patient produces fewer words, and his/her accompanying person plays a more important role.

Considering that C-ORAL-ESQ documents real consultations in which every participant is expected to behave naturally and spontaneously, this kind of variation was actually expected to be found in the corpus. The marked variation in patient profiles (regarding their sociodemographic features—see Figs. 3 and 4—as well as their clinical symptoms), the stage of their treatment, and their therapeutic needs in the day in which the recording took place, among many other factors, lead physicians, patients and patients' accompanying persons to behave in very specific ways that cannot be fully predicted. Because of that, the great variation regarding number of words and length of consultations could actually be considered as correlated to the spontaneous speech produced in this type of situation.

The variability among consultations can be seen with more details in the next two figures. Figure 1 shows a graph with the percentage of words produced by patients in relation to the words produced by other participants (health professionals, accompanying persons and intervenient participants) in each recording. As can be seen in the figure, in 17 recordings, patients produced at least 50% of the total number of words in the consultation (the highest case being 89% of words). In the following 13 recordings, patients produced less than 50% of the total number of words (the lowest case being 5% of words).

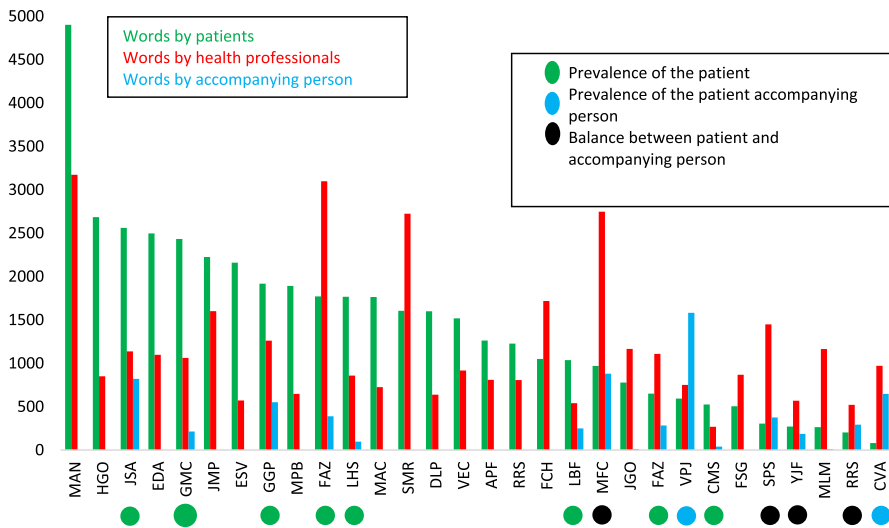


Fig. 2 Number of words produced by patients (green), health professionals (red) and accompanying persons (blue) in each recording. The colored dots indicate the prevalence of the patient over the accompanying person (green), the prevalence of the accompanying person over the patient (blue) or the balance between the two of them (black) in the recordings in which a patient's accompanying person participates. (Color figure online)

The dashed line separates the recordings in which the patient talks more than the other participants (yellow columns) from the recordings in which the opposite occurs (blue columns).

Figure 2 displays a grouped bar graph indicating the number of words produced by patients, health professionals and accompanying persons in each recording. The graph is ordered according to the number of words by patients, going from maximum (circa 5,000) to minimum (approximately 80). It can be noticed that, apart from a few exceptions (MAN, FAZ, SMR and MFC), the number of words produced by the health professionals is somehow constant, leading to the hypothesis that they play a similar role in most of the consultations. Considering this, it seems that, in the registrations with fewer words by the patient, the patient's accompanying person could play a more relevant role in the consultation. In fact, there are in total 14 consultations in which a patient's accompanying person is present (marked with colored dots below the initials). Among those recordings, in the ones in which the patient produces more than 1,500 words, the patient's accompanying person does not play an important role in terms of number of produced words (recordings marked with green dots). When the patient produces less than 1,500 words, on the other hand, there can be a balance between the words of the patient and those of his/her accompanying person (black dots) or even the prevalence of the number of words of the accompanying person (blue dots).

This type of information can be very relevant from a pragmatic point of view, since it allows for the understanding of some fundamental aspects of the interaction. These differences seem to correlate firstly to the patients' verbal behavior. This information can help to group patients in term of an important variable, and see

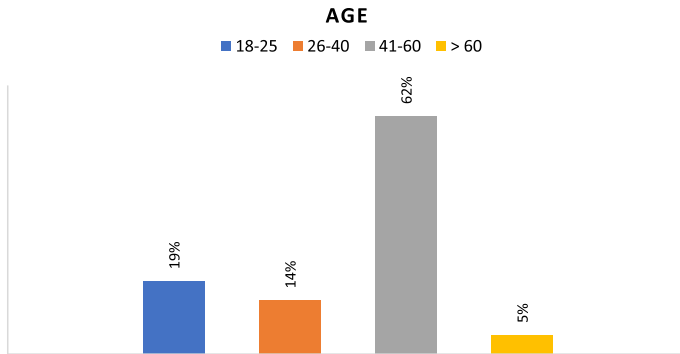
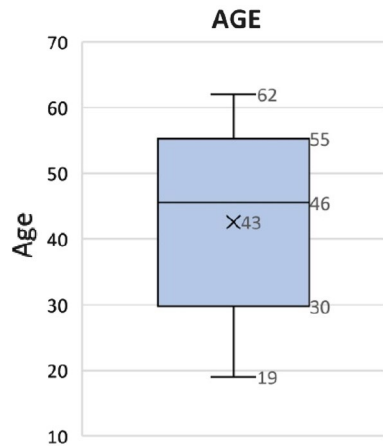


Fig. 3 Distribution of participants according to their age range, based on the adopted age groupings

Fig. 4 Distribution of participants according to their age



if this correlates with some kinds of pathologic characteristics (Cangemi et al., in press).

3.2 Sociolinguistic information

In this section, we present and comment the sociolinguistic data. We followed the same ranges adopted in the C-ORAL-BRASIL corpus (Raso & Mello, 2012; for a description of the corpus, see Raso, 2012; Mello, 2014; Raso & Mello, 2014).

Firstly, we need to say that all the patients are from the state of Minas Gerais, Brazil. In fact, IRS is the reference public psychiatric facility for the whole state. Secondly, IRS patients are male in their great majority. Consequently, the participants in the corpus so far are 28 males (75%) and only 8 females (25%). Figures 3 and 4 show age distribution, while Fig. 5 presents schooling data.

The patients are separated in four ranges of age. These specific ranges are the same adopted in the C-ORAL-BRASIL corpus. The more the two corpora

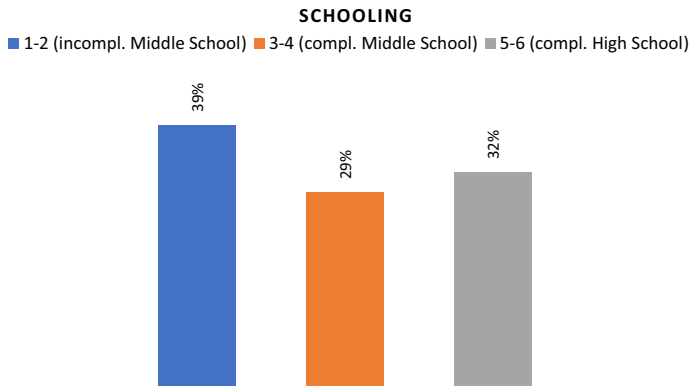


Fig. 5 Distribution of participants according to their schooling range

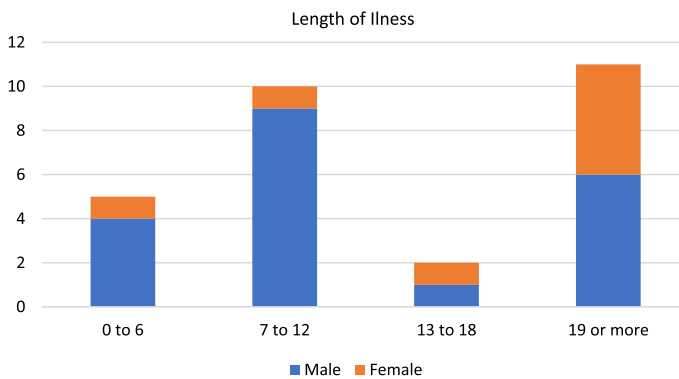


Fig. 6 Length of illness distribution

could be comparable, the better was for our scientific goals. As can be verified in Fig. 3, more than 60% of the participants pertain to age range C, between 41 and 60 years old, with a much lower presence of the other age ranges. Figure 4 offers a different visualization of the age range. As can be verified in Fig. 4, the age of the participants varies from 19 to 62 years old, with a median of 46 years old (SD 13,4).

Schooling distribution covers lower levels of education (Fig. 5). Among all patients, 39% did not complete the first study cycle, which means that they attended school for less than 8 years; 32% completed the first 8 years cycle (Middle School), and only 29% completed High School. This distribution is of course linked to the patients' professional profile, that shows that participants work (or used to work, since most of them are retired) in menial jobs or are housewives. This state of affairs is due to the fact that IRS is a public facility and that in Brazil middle and high socioeconomic classes usually have access to private health care and are not tended to in public hospitals.

Fig. 7 Length of illness distribution

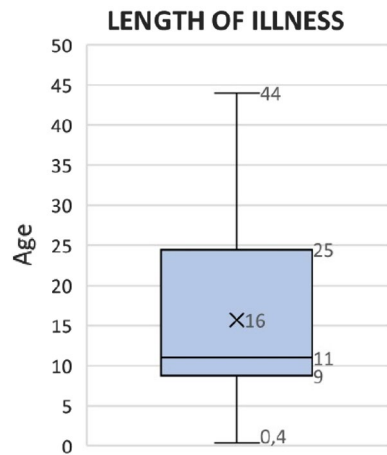


Figure 6 shows length of the illness and its distribution between men and women. Only 18% of the participants were diagnosed within the previous 7 years (actually, they were all diagnosed no more than 3 years before the recording), while 36% were diagnosed between 7 and 12 years before the recording, 7% between 13 and 16 years before the recording (all of them between 16 and 18 years before the recording) and 40% were diagnosed at least 19 years before the recording. Therefore, the participants constitute a population that has dealt with a long-term illness.

Figure 7 shows that the length of the illness varies from 0,4 to 44 years. The median is 11 years (SD 12), which means that 50% of all participants were diagnosed more than 11 years before the recording.

3.3 Setting, recording equipment and workflow

All the recordings were undertaken at IRS. The consultations recorded are all part of the ongoing treatment to which the patients are submitted. They are conducted by the same psychiatrist that takes care of each patient, in the same ambulatory in which they usually occur. The resident doctors are previously explicitly instructed to conduct the consultation as they usually would: they are not supposed to pressure the patient to talk more, less or differently. The residents do not have to follow any kind of plot determined by the C-ORAL-ESQ team whatsoever and are free to conduct the consultation according to their usual practice.

Consultations are carried privately, with the presence of doctors, patients and accompanying persons, without the presence of the recording team. This is possible because the C-ORAL-ESQ team make use of high-quality wireless equipment: a Tascam DR-100MKII digital recorder (configured to produce 24bit 48.0 k wave files) and two Sennheiser wireless clip-on microphone sets (each one composed by an SK100 G3 transmitter, an EK100 G3 receiver and an ME2 microphone). The recordings are carried in stereo channel.

The workflow of the recordings goes as follows. Once a resident doctor identifies a patient that potentially fits the C-ORAL-ESQ profile, he/she inquiries from the patient, during a consultation, whether his/her next session can be recorded in order to be used in scientific research. Whenever a patient gives this first and informal permission, the recording team is contacted by the resident to record the next session.

The recording team arrives at IRS at the scheduled day and sets the equipment. Before the consultation starts, the recording team briefly explains to the patient and the accompanying person the research aims and then reads the consent form to every person that will participate in the recording (the physician, the patient and the patient's accompanying person, when the latter is present at the consultation). After reading it, the team asks every participant to sign their respective consent form in case they agree to its terms (which were previously approved by both IRS and UFMG Ethics Committees). Finally, the recording team positions the microphone sets on the participants, leaves the consultation room and moves into another room in which they operate the digital recorder. The recording team goes back to the consultation room only to collect the equipment, after the consultation has ended.

As of recently, the Ethics Committees have approved an amendment to the C-ORAL-ESQ recording protocols that will make video recordings possible, in addition to audio recordings, during consultations. This will allow the study of co-speech gestures of hands and arms, and facial expressions of patients. Of course, this will be done through a process of anonymization that recovers different points of the face, and the movement of head and shoulders, in order to study facial expressions. Gestures and facial expressions will be aligned and tagged through ELAN.

3.4 Transcription and segmentation criteria

The complete framework for the transcription criteria can be found in Mello et al. (2012). These criteria were defined for the compilation of the C-ORAL-BRASIL corpus (Raso & Mello, 2012). The criteria were planned to aim at the following goals: (i) to provide easily readable texts; (ii) to capture phenomena ongoing possible grammaticalization or lexicalization processes; and (iii) to avoid difficulties for automatic PoS tagging. They were maintained for this corpus for two main reasons: (i) they allow research about morphosyntax and lexicon; (ii) they make it easier to compare results from the two corpora and to use the same parser, namely, PALAVRAS by Bick (2000) (see also Bick, 2012 and 2014).

Brazilian Portuguese spontaneous speech presents many phenomena of ongoing grammaticalization or lexicalization processes. Some examples are: the reduction of the pronominal forms, especially, but not only, personal pronouns (*você* > *cê*; *eles* > *es*; *aqueles* > *akes*); many apheretic lexical forms (*estar* > *tar*; *embora* > *bora*; *obrigado* > *brigado*); the reduction or neutralization of the verbal forms (*você/ele/nós/a gente/ vocês/ eles faz*), that tend to lose their person and number marks; the tendency to mark plural only on the determinant in noun phrases (*os meninos bonitos* > *os menino bonito*); the tendency to reduce the desinence of diminutives (*bonzinho* > *bonzim*), and many other less frequent phenomena.

The attention to these phenomena allows many studies on morphosyntactic and lexical changes that are currently occurring in spontaneous speech. The transcription accuracy of both orthographic and non-orthographic forms, as well as of criteria to transcribe interjections, acronyms, numerals or other forms, will be statistically validated. No specific phenomenon can have more than a 5% transcription error rate (otherwise the corpus undergoes a new revision and is tested again), but the experience due to the C-ORAL BRASIL I (Raso and Mello, 2012) and the C-ORAL-BRASIL II (Raso et al., forthcoming) shows that, following the reported process, the error rate is much smaller, especially for high frequency phenomena. This validation will be done at the end of the corpus compilation process, when all the needed elements will be acquired.

In Mello et al. (2012) also the segmentation criteria are presented and explained. These criteria are very important for a full understanding of the C-ORAL-ESQ corpus. In fact, the corpus is prosodically annotated in order to show the prosodic terminated units (TU) and their internal organization in intonation units (IU). The TU is defined as a minimal sequence of speech that exhibits both pragmatic and prosodic autonomy. This means that TUs must exhibit at least one illocution (which allows the pragmatic interpretation) and must end with a prosodic boundary that conveys the perception of prosodic conclusion. Therefore, we have two main levels of prosodic annotation: the TU and the IU. The intonation unit is defined on the basis of its perceptual prosodic correlate: the flow of speech is perceived in a prosodic envelope (Chafe, 1994; Crystal, 1975; Moneglia & Cresti, 1997), whose boundaries are two major prosodic breaks. The intonation units coincide with this envelope. Each TU is marked with two slashes at the end (//), while the boundary of an IU is marked by one slash (/). Of course, especially when the level of interaction is higher, many IUs can coincide with a TU. The TU is also the alignment unit in the software WinPitch and ELAN. Besides these two main annotations, two other important annotation symbols are used: the symbol (+) for interruptions (no matter the cause of them) and the symbol ([/n]) for retractings; in the latter case, the number (n) indicates the quantity of retracted words. It is also useful to say that usually filled pauses, that we call Time Takings, are isolated in an IU. Time Takings are always transcribed as (&he).

This kind of prosodic annotation follows almost the same criteria already used for the C-ORAL-ROM corpus (Cresti & Moneglia, 2005). Similar criteria are also used in the Santa Barbara corpus (Du Bois et al. 2000–2005), the CorpAfroAs (Mettouchi et al., 2015), the COSIH (Izre'el et al., 2001) and other spoken corpora.

The corpus is transcribed and prosodically annotated by linguistics students (usually undergraduate students) who have gone through a training period that normally lasts one semester.⁴ When they seem to be ready for the task, they take a Kappa test, and if they reach an agreement of at least 0.8 in the Kappa test (Fleiss, 1971) with one or more senior annotators, they are considered experts and can start to transcribe and segment the texts. Each transcription and segmentation is then revised at

⁴ The students receive a fellowship that includes, as part of their work, the transcription, revision and alignment process.

least twice, by different students. The revisors are individuals who have reached the highest Kappa score.

4 The informational annotation criteria and some pilot studies

Part of the corpus is informationally annotated according to L-AcT (see Cresti, 2000; Cresti & Moneglia, 2010; Moneglia & Raso, 2014; Cavalcante, 2020). Our goal is to informationally annotate at least 20 texts and 30,000 words uttered by patients, in order to study how individuals with schizophrenia manage the relation between prosody and information structure. We cannot fully explain the theory here, but we will try to present some important aspects that, in our view, make it very useful for the study of information structure and prosody, which is its main formal cue.

To observe how patients manage information structure in their discourse provides important inferences about cognitive aspects. This is especially true within the L-AcT paradigm, as will be better explained below, when more information about Topic and Appendix units will be presented.

The theory, originated by several decades of corpus driven work, is an extension of Austin's (1962) Speech Act Theory and individuates two reference units for speech (see Izre'el et al., 2020, and especially the chapters by Cresti, Raso et al. and Bossaglia et al.), i.e. two kinds of TUs: the *utterance* and the *stanza*.

The utterance is an information pattern, whose nucleus is an IU that hosts the illocution, which is seen as the core of any communication activity, since it conveys an action toward the interlocutor, and is therefore communicatively interpretable. This IU is called *Comment*. The utterance can be simple or compound. It is simple if it is formed just by the Comment, that is, just by the illocutionary IU. It is compound if, besides the Comment, one or more non-illocutionary units are patterned with the illocution, to which they are functionally and prosodically subordinated.

The stanza is made up by two or more juxtaposed patterns. Each pattern can be formed by one illocutionary unit or by it and one or more non-illocutionary ones. However, differently from the utterance, the pattern is not concluded (except the last one), but ends with a prosodic continuity signal followed by the performance of another pattern. Stanzas, therefore, are built by two or more (sometimes many more) juxtaposed informational patterns, linked by a continuation prosodic signal. Stanzas are very frequent in monologues and, in general, in speech with low interactive level. While utterances feature a great variety of illocutions, stanzas tend to be much more homogeneous in their illocutionary value.

As pointed out, we cannot account for a full description of the theory here, but we have to say something about a few non-illocutionary units that have already received some attention in studies within the C-ORAL-ESQ project and by other studies using interactions between Italian individuals with schizophrenia and their physician (see Dovetto & Gemelli, 2012; Dovetto et al., 2015).

Firstly, we need to say that the non-illocutionary intonation/information units can be grouped in two major types: those units that build the semantic text of the TU and those that are directed to the interlocutor, serving to regulate the communication. The latter ones are normally known as Discourse Markers in different paradigms

(Fischer, 2006; Raso, 2014; Raso et al., 2022). Each information unit is defined based on its specific pragmatic function, its prosodic form and its distribution with respect to the Comment unit.

Among the non-illocutionary units, some are well known in the literature, as the Parenthetical (Dehé, 2014; Dehé & Kavalova, 2007; Schneider, 2007 and 2014; Tucci, 2004 and 2009), others, as the Topic and the Appendix, are defined differently with respect to mainstream approaches (see Krifka & Musan, 2012, and, for our approach, Raso et al., 2017, Cavalcante, 2020, and Cavalcante and al., in press), and others yet, like the Locutive Introducer, are not treated in other frameworks, as far as we know.

The studies carried on so far within this theoretical framework on texts produced by individuals with schizophrenia led to some initial interesting results. Therefore, we will concentrate on the units that were investigated in these first studies. Dovetto et al., (2015) made some qualitative observations on three patients of an Italian corpus. They observed that schizophrenic patients produce less illocutionary variation than non-pathological speakers. They also observed a reduction of the production of Topic and an increase of production of Appendix. These results were confirmed by Costa (2022) on quantitative data of the C-ORAL-ESQ, using six tagged texts (approximately 9,000 words produced by patients). These findings are analyzed in detail in Rocha et al. (in press).

In order to better understand what this can mean at a cognitive level, it is important to say something more about Topic and Appendix information units. Topic is defined in L-Act as the field of application of the illocutionary force. This means that Topic provides the cognitive domain in which the illocution must be interpreted. If we do not have a Topic, the illocution is inferentially applied to some cognitive domain that is considered obvious for contextual reasons. However, if we want to change this cognitive domain or if, for any reason, we need or want to establish it or make it clear or make it explicit, we need to use a Topic information unit. From a cognitive point of view, this means that the speaker needs to understand if at a certain point of the interaction it is appropriate to signal to the hearer the cognitive domain of application of the illocutionary force still to be performed, in order to situate the hearer. Besides this, of course, the speaker needs to perform the Topic in the appropriate segmental and prosodic way.

The prosody of Topic is rather complex. Previous studies (Cavalcante, 2020; Raso et al., 2017, and Cavalcante et al., in press), analyzing data from corpora of English, Italian, and Brazilian and European Portuguese, have shown that, in all these languages, three different prosodic forms of the Topic unit can be found. The frequency of each form seems to correlate with rhythmic properties of the language and not with specific subfunctions. In any case, the three forms feature a syllabic nucleus of one or two syllables (in one form the nucleus can reach a maximum of four syllables) where an informationally functional prominence is placed. This means that the prosodic parameters of the specific prominence of the nucleus carry the informational function of Topic, while the rest of the unit simply carries the semantic content, but does not have any influence on the conveyed information value. Regardless of the specific form, the prominence that conveys the Topic function always features at least one (sometimes two) f0 movement on specific syllables, together with

an increase of duration and intensity. This means that the speaker needs to manage the correct prosodic parameters, distributed in a short length of time.

Considering all these aspects, we can say that the Topic information unit is especially complex both from a cognitive and a prosodic point of view. As mentioned, previous studies found that schizophrenic speech tends to use less Topics than non-pathological speech. Costa (2022) shows that this decrease is statistically significant.

On the other hand, the studies on the Italian data observe an increase of Appendix information units in schizophrenic speech. This was confirmed by Costa (2022) for Brazilian Portuguese speakers-patients recorded in C-ORAL-ESQ. However, Costa did not find statistical significance in this increase. Nevertheless, it is worthwhile to describe this unit, since it features characteristics to some extent opposite to those of the Topic unit. In fact, Appendix can be described as a textual integration of the Comment unit with cognitively already given information, such as the recovery of textual or contextual information, or simply a repetition or a semantically empty word. Interestingly, Appendix has a flat or falling profile, without any prominence; this unit is also short, thus far from the lexical or syntactic complexity that Topic may reach. Therefore, Appendix is a very simple information unit both from a cognitive and a prosodic point of view.

A third interesting result of the studies on the Italian data, as previously mentioned, is the reduction of illocutionary variability. Before delving into this issue, we need to explain the methodology that was used to compare schizophrenic and non-pathologic speech, since neither the Italian researchers nor the Brazilian ones had the availability of a real control corpus to carry on their comparative work. While the Italian scholars used non-pathological interviews extracted from the C-ORAL-ROM to compare to the information structure of the pathological speakers, the Brazilian team developed a methodology that is described in detail in Rocha et al. (in press) and that we will briefly summarize here.

Considering that C-ORAL-ESQ could not count on a control corpus, for our first studies we had to conceive a methodology that could allow C-ORAL-BRASIL I to become to a certain extent the control corpus of C-ORAL-ESQ. C-ORAL-BRASIL features dialogues, monologues and conversations in many different situations, but not in situations comparable to consultation interactions between patients and physicians. Therefore, we did not compare texts, but structures. More precisely, we chose the same number of stanzas with the same number of patterns, and compared the complexity of the patterns. For instance, we compared stanzas with two illocutionary patterns from C-ORAL-ESQ with stanzas with two illocutionary patterns from C-ORAL-BRASIL I, we compared stanzas with three patterns of one corpus to stanzas with three patterns in the other corpus, and so on, until reaching stanzas with six patterns. What emerged is that the textual complexity of stanzas of the C-ORAL-BRASIL data is significantly higher than that of the C-ORAL-ESQ. This means that, given the same number of patterns, non-pathological speech shows a much higher degree of complexity, mainly due to a higher presence of all types of textual units (with the significative exception of Appendixes) and especially of Topics.

Since our comparison was concentrated on stanzas, in principle it was not possible to analyze the illocutionary variation. As we said, stanzas are characterized by a high degree of illocutionary homogeneity, since they appear where the

interactivity is much lower and tend to be part of narrations, argumentations or in general moments in which the speaker conducts a flux of thinking, as Chafe (1994) has already observed by analyzing the Pear stories. Nevertheless, there is a specific illocutionary unit that allows us to confirm the impression of the Italian scholars.

This specific illocutionary unit is the *Multiple Comment* (CMM). Multiple Comments are not a single prosodic unit, but a pattern of two or (rarely) more prosodic units that have illocutionary value by themselves but are interpreted in a holistic way, as a single nucleus of the pattern, formed by more patterned illocutions. Some examples of these patterned illocutions are lists, comparisons, reinforcements (the repetition of the same illocution in order to reinforce it, like *Yes / I agree //*), functional callings (that is, calling someone in order to attract his/her attention to a different illocution like *John / thank you //* or *John / be careful //*). Multiple Comments can be found both in utterances and in stanzas. What we observed, analyzing stanzas, is that the frequency of CMMs in non-pathological speech is higher than in schizophrenic speech in a statistically significant degree. This finding can be interpreted as an indirect confirmation of what was observed studying the Italian data, i.e., a higher illocutionary variation.

Summing up these results, we can hypothesize that schizophrenic speech is both cognitively and prosodically simpler than non-pathological speech. The issue at this point is related to how to interpret these results. Possible questions would be: do schizophrenic patients have problems related to the prosodic component and this, in turn, causes difficulty in performing information units with more complex prosodic structure? On the other hand, do they have some kind of cognitive difficulty that leads to a simpler information structure that affects to a greater extent those units that are cognitively more complex and, at the same time, more complex in their prosodic performance? Of course, one difficulty does not necessarily exclude the other one.

There are some studies available about prosody in schizophrenic speech. Usually these studies (like Cohen et al., 2014; Alpert et al., 2000; Covington et al., 2012; Cohen et al., 2013; Compton et al., 2018) compare the prosodic variation between schizophrenic and non-pathological speakers in large stretches of speech. In general, these studies have found that schizophrenic speech features less prosodic variation, but the results are usually below the level of statistical significance. One interesting study (Martínez-Sánchez et al., 2015) shows that the difference between the prosodic variation of the two groups becomes clearly significant if what is measured is not the mean variation of large stretches but the intrasyllabic variation. This could be related to our findings, because both Topics and illocutionary units are characterized by functional prominences that are usually located in one single syllable. If schizophrenic speakers have a significantly lower degree of intrasyllabic prosodic variation, this could be immediately related to their difficulty in performing Topics and illocutionary variation. It would be very interesting to check this hypothesis measuring what happens in the performance of the prosodic nuclei of both Topics and Comments, the only textual information units that feature a functional prosodic prominence.

During the informational tagging of the texts of C-ORAL-ESQ, we frequently had the impression that units that functionally work as Topic units are performed

with a non-satisfactory prosodic prominence, which may lead to difficulty in recognizing the function of the unit. One of our next steps will be to measure, with an appropriate software, the prominences of both pathological and non-pathological speech, in order to verify the cited study on a large amount of data.

5 Conclusion

We have presented a project that focuses on studies about the spontaneous speech of individuals with schizophrenia. The project is the outcome of the collaboration between the linguistics team at the Lab for Empirical and Experimental Studies of Language (LEEL), at the Federal University of Minas Gerais, and the ESQUIZO research group, formed by psychiatrists at the same university and at IRS. The project bases its studies on a spontaneous speech corpus of consulting sessions between psychiatrists and patients in a Brazilian public psychiatric facility. In this paper, we described the medical protocols and the methodology for the corpus compilation. We also presented the results of some pilot studies about information structure and prosody, and traced some possible future lines of investigation.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by TR, HM and BR. The first draft of the manuscript was written by TR, BR, JVS, BFC and LMM; and all authors commented on previous versions of the manuscript. TR and HM revised the final version of the manuscript. All authors read and approved the final manuscript.

Funding Partial financial support for the research reported was received from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

Data availability The datasets generated and analyzed during the current study are not publicly available because the corpus is still being compiled, but are available from the corresponding author on reasonable requests.

Declarations

Conflict of interest There are no conflicts of interest to declare.

References

- Alpert, M., Rosenberg, S. D., Pouget, E. R., & Shaw, R. J. (2000). Prosody and lexical accuracy in flat affect schizophrenia. *Psychiatry Research*, 97(2–3), 107–118. [https://doi.org/10.1016/S0165-1781\(00\)00231-6](https://doi.org/10.1016/S0165-1781(00)00231-6)
- Austin, J. L. (1962). *How to do things with words*. Cambridge University Press.
- Barnes, T. R. (1989). A rating scale for drug-induced akathisia. *British Journal of Psychiatry*, 154, 672–676.
- Bick, E. (2000). *The parsing system “PALAVRAS”: Automatic grammatical analysis of Portuguese in a constraint grammar framework*. Aarhus University Press.

- Bick, E. (2012). A anotação gramatical do C-ORAL-BRASIL. In T. Raso & H. Mello (Eds.), *O Corpus C-ORAL-BRASIL I Corpus de referência do português falado informal* (pp. 223–267). UFMG Press.
- Bick, E. (2014). The grammatical annotation of speech corpora: Techniques and perspectives. In T. Raso & H. Mello (Eds.), *Spoken Corpora and Linguistic Studies* (Vol. 61, pp. 105–128). John Benjamins.
- Bryson, G., Bell, M., & Lysaker, P. (1997). Affect recognition in schizophrenia: a function of global impairment or a specific cognitive deficit. *Psychiatry Research*, 71, 105–113.
- Cangemi, F., Grice, M., Janz, A., Lucarini, V., Spaniol, M. & Vogeley, K., (In press, in this volume). Content-free speech activity records: interviews with people with schizophrenia.
- Cavalcante, F. (2020). The information unit of Topic: a crosslinguistic, statistical study based on spontaneous speech corpora. Ph.D. Dissertation. Belo Horizonte: UFMG.
- Cavalcante, F., Raso, T., Barbosa, P. (In press). The Topic information unit: modeling prosodic forms in a crosslinguistic perspective. In: Marina Castagneto, Miriam Ravetto (eds), *La comunicazione parlata*.
- Chafe, W. (1994). *Discourse, Consciousness and Time*. Chicago & London: The University of Chicago Press.
- Cohen, A. S., Kim, Y., & Najolia, G. M. (2013). Psychiatric symptom versus neurocognitive correlates of diminished expressivity in schizophrenia and mood disorders. *Schizophrenia Research*, 146(1–3), 249–253.
- Cohen, A. S., Mitchell, K. R., & Ellevåg, B. (2014). What do we really know about blunted vocal affect and alolia? A meta-analysis of objective assessments. *Schizophrenia Research*, 159, 533–538. <https://doi.org/10.1016/j.schres.2014.09.013>
- Compton, M. T., Lunden, A., Cleary, S. D., Pauselli, L., Alolayan, Y., Halpern, B., Broussard, B., Crisafio, A., Capulong, L., Balducci, P. M., Bernardini, F., & Covington, M. A. (2018). The aprosody of schizophrenia: Computationally derived acoustic phonetic underpinnings of monotone speech. *Schizophrenia Research*, 197, 392–399. <https://doi.org/10.1016/j.schres.2018.01.007>
- Corcoran, R., Mercer, G., & Frith, C. D. (1995). Schizophrenia, symptomatology and social inference: Investigating “theory of mind” in people with schizophrenia. *Schizophrenia Research*, 17(1), 5–13. [https://doi.org/10.1016/0920-9964\(95\)00024-g](https://doi.org/10.1016/0920-9964(95)00024-g). PMID: 8541250.
- Costa Jr., J. C. (2022). Padrão informacional de stanzas de pacientes com esquizofrenia. PhD thesis, Universidade Federal
- Covington, M. A., Lunden, S. L., Cristofaro, S. L., Wan, C. R., Bailey, C. T., Broussard, B., Fogarty, R., Johnson, S., Zhang, S., & Compton, M. T. (2012). Phonetic measures of reduced tongue movement correlate with negative symptom severity in hospitalized patients with first-episode schizophrenia-spectrum disorders. *Schizophrenia Research*, 142(1–3), 93–95. <https://doi.org/10.1016/j.schres.2012.10.005>
- Cresti, E. (2000). Corpus di italiano parlato. Firenze: Accademia della Crusca. In E. Cresti & M. Moneglia (Eds.), *C-ORAL-ROM: integrated reference corpora for spoken Romance languages* (pp. 40–51). John Benjamins.
- Cresti, E. (2005). Notes on lexical strategy, structural strategies and surface clause indexes in the C-ORAL-ROM spoken corpora. In E. Cresti & M. Moneglia (Eds.), *C-ORAL-ROM: Integrated reference corpora for spoken Romance Languages* (pp. 209–256). John Benjamins.
- Crystal, D. (1975). *The English tone of voice*. Edward Arnold.
- Cresti, E.; Moneglia, M. (2010) Informational patterning theory and the corpus-based description of spoken language: The compositionality issue in the topic-comment pattern. In: Moneglia, M.; Panunzi, A. (Eds.). *Bootstrapping information from corpora in a crossLinguistic perspective*. Firenze: Firenze University Press, 2010, 13–45. Disponível em: <http://www.oapen.org/search?identifier=343705>. Acessado em: 02/11/2018
- Dehé, N. (2014). *Parentheticals in Spoken English: The Syntax-Prosody relation*. Cambridge University Press.
- Dehé, N. & Kavalova, Y. (2007) *Parentheticals*. Linguistics Today, 106. Amsterdam/Philadelphia: John Benjamins.
- Dovetto, F.M., Cresti, E. & Rocha, B (eds). Schizofrenia tra prosodia e lessico. Prime analisi. *Studi Italiani di Linguistica Teorica e Applicata*. 44(3): 486–507
- Dovetto, F. M., Cresti, E., & Rocha, B. (2015). Schizofrenia tra prosodia e lessico Prime analisi. *Studi Italiani di Linguistica Teorica e Applicata*, 44(3), 486–507.
- Dovetto, F.M. & Gemelli, M. (Eds.) (2012). *Il parlare matto. Schizofrenia tra fenomenologia e linguistica. Il corpus CIPPS*. Roma: Aracne

- Du Bois, J. W.; Chafe, W. L.; Meyer, C.; Thompson, S. A. & Martey, N. (2000–2005) Santa Barbara Corpus of Spoken American English. CD-ROM. Philadelphia: Linguistic Data Consortium.
- ELAN (2022) (Version 6.4) [Computer software. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from <https://archive.mpi.nl/tla/ela>
- First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2016). Structured clinical interview for DSM-5 disorders, clinician version (SCID-5-CV) (1st eds). American Psychiatric Association.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378–382.
- Fischer, K. (Ed.). (2006). *Approaches to discourse particles*. Bingley.
- Forj6, M., Neto, D., Abad, A., Pinto, H. S. & Gago, J. A new European Portuguese corpus for the study of Psychosis through speech analysis. In: Proceedings of the 13th Conference of Language Resources and Evaluation (LREC). European Language Resources Association (ELRA), p. 7298–7304.
- Galderisi, S., Mucci, A., Buchanan, R. W., & Arango, C. (2018). Negative symptoms of schizophrenia: New developments and unanswered research questions. *The Lancet Psychiatry*, 5(8), 664–677.
- Hoekert, M., & Aleman, A. (2007). Impaired recognition and expression of emotional prosody in schizophrenia: Review and meta-analysis. *Schizophrenia Research*, 96, 135–145.
- Izre'el, S. H., Hary, B., & Rahav, G. (2001). Designing COSIH: The corpus of spoken Israeli Hebrew. *International Journal of Corpus Linguistics*, 6(2), 171–197.
- Izre'el, Sh., Mello, H., Panunzi, P. & Raso, T. (Eds.) (2020). In search of basic units of spoken language: a corpus-driven approach. (Studies in Corpus Linguistics 94.). Amsterdam/Philadelphia: John Benjamins.
- Kay, S., Fiszbein, A., & Opler, L. (1987). The positive and negative syndrome scale (PANSS) for Schizophrenia. *Schizophrenia Bulletin*, 13(2), 261–276.
- Keefe, R. S., Poe, M., Walker, T. M., Kang, J. W., & Harvey, P. D. (2006). The Schizophrenia cognition rating scale: An interview-based assessment and its relationship to cognition, real-world functioning, and functional capacity. *The American Journal of Psychiatry*, 163(3), 426–432.
- Li, Y., Lin, Y., Ding, H., & Li, Ch. (2019). Speech databases for mental disorders: A systematic review. *Gen Psychiatry*, 32(3), e100043.
- Kahn, R. S., & Keefe, R. S. (2013). Schizophrenia is a cognitive illness: Time for a change in focus. *JAMA Psychiatry*, 70(10), 1107–1112.
- Krifka, M., & Musan, R. (Eds.). (2012). *The expression of information structure*. De Gruyter Mouton.
- Lam, M., Wang, M., Huang, W., Eng, G. K., Rapisarda, A., Kraus, M., Kang, S., Keefe, R. S. E., & Lee, J. (2017). Establishing the brief assessment of cognition - short form. *J Psychiatric Research*, 93, 1–11. <https://doi.org/10.1016/j.jpsychires.2017.05.006>
- Lieberman, J. A., Stroup, T. S., McEvoy, J. P., Swartz, M. S., Rosenheck, R. A., Perkins, D. O., Keefe, R. S. E., Davis, S. M., Davis, C. E., Lebowitz, B. D., Severe, J., Hsiao, J. K., & Severe, J. (2005). Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *New England Journal of Medicine*, 353(12), 1209–1223.
- Louzã Neto MR Escalas de avaliação de efeitos colaterais de antipsicóticos (1998). *Revista de Psiquiatria Clínica*. 25 (6): 357–367
- Martin, Ph. (2015). WinPitch. www.winpitch.com
- Martínez-Sánchez, F., Muela-Martínez, J., Cortés-Soto, P., Meilán, J., Ferrándiz, J., Egea-Caparrós, D., & Valverde, I. M. (2015). Can the acoustic analysis of expressive prosody discriminate schizophrenia? *The Spanish Journal of Psychology*, 18, 1–9. <https://doi.org/10.1017/sjp.2015.85>
- Mausbach, B. T., Harvey, P. D., Goldman, S. R., Jeste, D. V., & Patterson, T. L. (2007). Development of a brief scale of everyday functioning in persons with serious mental illness. *Schizophrenia Bulletin*, 33(6), 1364–1372.
- Mello, H. (2014). Methodological issues for spontaneous speech corpora compilation: The case of C-ORAL-BRASIL. In T. Raso & H. Mello (Eds.), *Spoken corpora and linguistic studies*. John Benjamins.
- Mello, H.; Raso, T.; Mittmann, M. M.; Vale, H.; Côrtes, P. (2012) Transcrição e segmentação prosódica do corpus c-oral-brasil: critérios de implementação e validação. In: Raso, T.; Mello, H. (Eds). C-ORAL-BRASIL I. Corpus de referência do português brasileiro falado informal. Belo Horizonte: UFMG: 125–174.
- Mettouchi, A., Vanhove, M., & Caubet, D. (2015) Corpus-based Studies of Lesser-described Languages: The CorpAfroAs corpus of spoken AfroAsiatic languages. In: Mettouchi, A., Vanhove, M., & Caubet, D. (Eds). John Benjamins, Amsterdam
- Michelas, A., Faget, C., Portes, C., Lienhart, A. S., Boyer, L., Lançon, C., & Champagne-Lavau, M. (2014). Do patients with schizophrenia use prosody to encode contrastive discourse status? *Frontiers in Psychology*, 18(5), 755. <https://doi.org/10.3389/fpsyg.2014.00755>. PMID:25101025;PMCID:PMC4102879

- Moneglia, M., & Cresti, E. (1997). L'intonazione e i criteri di trascrizione del parlato adulto e infantile. In U. Bortolini & E. Pizzuto (Eds.), *Il Progetto CHILDES Italia* (pp. 57–90). Del Cerro.
- Moneglia, M., & Raso, T. (2014). Notes on language into act theory. In T. Raso & H. Mello (Eds.), *Spoken Corpora and Linguistic Studies* (pp. 468–495). John Benjamins.
- Morosini, P. L., Magliano, L., Brambilla, L., Ugolini, S., & Pioli, R. (2000). Development, reliability and acceptability of a new version of the DSM-IV Social and Occupational Functioning Assessment Scale (SOFAS) to assess routine social functioning. *Acta Psychiatrica Scandinavica*, 101(4), 323–329.
- Rapcan, V., D'Arcy, S., Yeap, S., Afzal, N., Thakore, J., & Reilly, R. B. (2010). Acoustic and temporal analysis of speech: A potential biomarker for schizophrenia. *Medical Engineering & Physics*, 32(9), 1074–1079.
- Raso, T. (2012). O corpus C-ORAL-BRASIL. In : Raso, T. & Mello, H. C-ORAL-BRASIL I. Corpus de referência do português falado informal. Belo Horizonte: UFMG: 55–90.
- Raso, T. (2014). Prosodic constraints for discourse markers. In T. Raso & H. Mello (Eds.), *Spoken Corpora and Linguistic Studies* (pp. 411–467). John Benjamins. <https://doi.org/10.1075/scl.61.14ras>
- Raso, T.; Cavalcante, F.; Mittmann, M. (2017). Prosodic forms of the Topic information unit in a cross-linguistic perspective: a first survey. De Meo, A. & Dovetto, M.F. (Eds.). Proceedings of the SLI-GSCP International Conference. Roma: Aracne editrice.
- Raso, T. & Mello, H. (Eds.) (2012). C-ORAL-BRASIL I. Corpus de referência do português falado informal. Belo Horizonte: UFMG.
- Raso, T. & Mello, H. (2014) C-ORAL-BRASIL: Description, Methodology and Theoretical Frame. In: Berber Sardinha, T. São Bento Ferreira, T.L (eds). Working with Portuguese Corpora. Bloomsbury, London. Pp. 257–276
- Raso, T., Mello, H. & Ferrari, L. (forthcoming). O C-ORAL-BRASIL II: Formal em contexto natural, mídia, telefone.
- Raso, T. & Mittmann, M. (2012). As principais medidas da fala. In: Raso, T. & Mello, H. (eds). C-ORAL-BRASIL I. Corpus de referência do português falado informal. Belo Horizonte: UFMG, p. 177–221.
- Raso, T., Rilliard, A., Santos, S., & Moraes, J. (2022). Modeling the prosodic forms of discourse markers. *Domínios da Linguagem*, 16(4), 1436–1488.
- Rocha, B., Raso, T. Mello, H & Ferrari, L. (in press). Information structure of individuals with schizophrenia: methodology and first analyses from corpus based data. In: Dovetto, M.F., Raso, T. & Sorianello, P. Le patologie del linguaggio: studi e risorse tra crossdisciplinarietà e interdisciplinarietà. Chimera: Romance Corpora and Linguistic Studies.
- Schneider, S. (2007). Reduced Parenthetical Clauses as Mitigators: A corpus study of spoken French, Italian and Spanish. Amsterdam/Philadelphia: John Benjamins.
- Schneider, S. (2014). Parenthesis: fundamental features, meanings, discourse functions and ellipsis. In: Marlies K.; OTT, D.; De Vries, M. (Eds.). Parenthesis and ellipsis: crosslinguistic and theoretical perspectives. Berlin, De Gruyter Mouton: 277–300.
- Simpson, G. M., & Angus, J. W. S. (1970). A rating scale for extrapyramidal side effects. *Acta Psychiatrica Scandinavica (suppl.)*, 212, 11–19.
- Tucci, I. (2004). L'inciso: caratteristiche morfosintattiche e intonative in un corpus di riferimento. In: Albano Leone, F., Cutugno, F., Pettorino, M., Savy, R. (Eds.). Il parlato italiano. Atti del Convegno Nazionale. Napoli: D'Auria Editore: 1–14.
- Tucci, I. (2009). Obiter dictum: La funzione informativa delle unità parentetiche. In: Pettorino, M., Gianni, A. & Dovetto, F. (Eds.). *La Comunicazione parlata 3*, Atti del GSCP, 3 vol.: 635–654, Napoli: Università l'Orientale Press.
- Witterburg, P., Brugman, H., Russel, A., Klassmann, A. & Sloetjes, H. ELAN: A Professional Framework for Multimodality Research. In: Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC). European Language Resources Association (ELRA), 2006: 1556–1559.
- World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, 310(20), 2191–2194.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.