Information structure in the speech of individuals with schizophrenia

Methodology and first analyses from complex structure of corpus based data

Bruno Rocha°, Tommaso Raso°, Heliana Mello°, Lucia Ferrari°
°Universidade Federal de Minas Gerais (UFMG)

This paper focuses on information structure in the speech of individuals with schizophrenia, according to the theoretical framework found in the *Language into Act Theory*. A specific textual structure, called *stanza*, and its informational complexity, have been compared using two corpora: the C-ORAL-ESQ for pathological speech, and the C-ORAL-BRASIL for non-pathological speech. Firstly, we briefly explain the theory and the concept of *stanza*. Then, we explain in depth the methodology adopted for comparing pathological and non-pathological speech, using the two corpora. Additionally, we discuss the results found, part of which have statistical significance. Finally, we propose a first explanation of the results, focusing on cognitive and prosodic aspects that could provide insights for further research.

Keywords: Individuals with schizophrenia; information structure; spontaneous speech; corpus; prosody

1. Introduction

This paper has three goals, which are: (i) to present a methodology for the study of the informational structure of the speech of individuals with schizophrenia; (2) to synthesize the most relevant results obtained in a first study using this methodology, and (3) to analyze the results in order to draw some conclusions that may lead to insights on cognitive and prosodic aspects of individuals with schizophrenia.

The methodology was designed at the LEEL lab unit at the Federal University of Minas Gerais (UFMG), as part of the C-ORAL-ESQ project, coordinated by

Tommaso Raso and João V. Salgado. This project aims to compile a corpus of spontaneous speech of Brazilian individuals with schizophrenia during their periodical interactions with physicians in a medical environment, and to study it for linguistic and medical purposes.

The methodology is designed to study the information structure of individuals with schizophrenia describing their information patterns with data provided by the C-ORAL-ESQ corpus (Rocha *et al.*, in preparation) and to compare them with the information patterns produced by individuals without mental disorders. Since the C-ORAL-ESQ project has not yet compiled a medic control corpus, the methodology was thought to make it possible to use data from a general speech corpus – the C-ORAL-BRASIL (Raso & Mello, 2012), a reference corpus of Brazilian Portuguese spontaneous speech. The goal is, therefore, to overcome the problems that normally result from the adoption of two non-comparable corpora for the study of a specific phenomenon.

The proposed solution, explained in detail in this paper, is to study the information structure of *stanzas* produced in monological excerpts. Stanzas (Cresti, 2009) are, along with utterances, the basic units of human communication, according to the *Language into Act Theory* (L-AcT; Cresti, 2000; Moneglia & Raso, 2014). They correspond to prosodically autonomous sequences that convey more than one speech act (Austin, 1962).

Section 2 will briefly present the C-ORAL-ESQ corpus. Section 3 will explain the main properties of the basic units of spoken language (utterances and stanzas), which are necessary to understand the proposed methodology itself along with the fundamentals of L-AcT. Then, section 4 presents the methodology for the study of information structure in a comparative key, taking data from the two corpora in consideration. Section 5 introduces the main results obtained from a computational and statistical analysis as reported in Costa (2022). Finally, section 6 analyzes Costa's data in order to draw some conclusions regarding the main strategies employed by individuals with schizophrenia to build information structure in their speech and what they may reveal about their cognition and prosody.

2. The C-ORAL-ESQ corpus

C-ORAL-ESQ is a spoken corpus aimed to document the spontaneous speech of Brazilian individuals with Schizophrenia. The corpus, which is being compiled, focuses on the speech produced during psychiatric consultations between patients and their psychiatrists in their regular treatment, carried out at Instituto Raul Soares (IRS – FHEMIG) – a public psychiatric facility in Belo Horizonte, Brazil. In

Brazil there is a marked social profile differentiation between people who seek treatment in public or in private facilities. This means that the economical and sociocultural stratum of the average IRS patient could be labeled low or very low.

The corpus, upon completion, will be comprised of at least 40 registrations, each with approximately 1.500 words produced by the patients (plus the words produced by their legal guardians and physicians). So far, the corpus features 30 recordings, 43.008 words produced by patients (mean 1.434, SD 1.023) and 42.436 words produced by other participants (physicians and legal guardians), resulting on a total of 85.444 words. The large variation in the number of words uttered by patients is an expected feature of the corpus. This is due to the fact that all registrations portray real consultations carried out during the treatment, and patients differ from one another on their symptomatology, besides their sociodemographic characteristics. For more details on the C-ORAL-ESQ, see Rocha *et al.* (2020) and Rocha (2019).

3. Basic communicative units of spoken language

3.1. The discussion about basic units in speech

There is much discussion in linguistics about the nature and properties of the basic communicative units of spoken language. Formalist approaches tend to conceive them as a syntactical unit, the sentence, defined as the major projection of a V (Jackendoff, 1977). Based on the primacy of syntax over other levels of linguistic analysis on various formalist approaches, prosody is often considered to be subordinated to syntactic structure, acting on the linguistic level by making explicit the position of syntactic boundaries, disambiguating structures, etc.

On the other hand, functionalist approaches tend to define the basic communicative units of spoken language based on pragmatic and/or cognitive properties. Syntax is not necessarily considered the main linguistic level where to look for the features that characterize a basic unit. This is due to the fact that the very reason for considering a linguistic unit as a basic discourse unit is its capacity of conveying a minimal communicative function, and not its structural characteristics. Various authors observe, in fact, that around one third of the utterances in spoken communication do not carry a verb (Biber et al., 1999 for English; Cresti, 2005 for some Romance Languages; Raso and Mittmann, 2012, for Brazilian Portuguese). Therefore, various frameworks claim that prosody is the primary linguistic marker responsible for setting the boundaries of the communicative units of speech – even if those boundaries may eventually coincide with syntactic ones – and for establishing other features that allow the perception of a linguistic

sequence as a communicative one, a speech act. For different proposals of basic units that share a general functional approach, see Izre'el *et al.* (2020).

In this article, we follow a functional approach, L-AcT, designed to analyze spontaneous speech from corpus data. According to it, the basic communicative units of spoken language are defined as the minimal stretches of speech that are prosodically and pragmatically interpretable in isolation. Their autonomy derives from the perception that a stretch of speech constitutes a terminated sequence that conveys (at least) one speech act (Austin, 1962). Each terminated sequence ends with a terminal prosodic boundary and can be formed by one or more prosodic/information units. In the latter case, the different units are separated by non-terminal boundaries.

The idea that spontaneous speech is segmented into smaller units primarily by perceivable prosodic boundaries is well spread especially among functionalists who work with spontaneous corpus data[1]. The strong reliability of prosodic segmentation is confirmed by the application of different rounds of the Kappa test (Fleiss, 1971), that easily show a level of agreement higher than 0.8. Among other resources, this can be verified in the data of C-ORAL-ROM corpora (Cresti and Moneglia, 2005), whose results are presented by Moneglia et al. (2005), and C-ORAL-BRASIL (Mello *et al*., 2012).

According to L-AcT, prosody is also responsible for carrying at least two other crucial functions for spoken communication: to identify which prosodic units of a terminated sequence convey an illocution (in contrast to non-illocutionary units) and (together with some pragmatic features) to signal the specific type of illocution conveyed by them. The same sequence of words can receive different segmentations and convey different speech acts, depending on their prosodic profile, as shown in example (1) (see Izre'el *et al*., 2020a)

(1)
a. People (Calling)! Give John the book I promised him (Order)!
b. People give John the book I promised him (Assertion).
c. People give John the book (Question)? I promised him (Assertion).
d. People (Calling)! Give John the book (Order)! I promised him (Assertion).

[1] See Panunzi et al. (2020) for a comparison among annotations that scholars subscribing to different theoretical frameworks made of the same two stretches extracted from the Santa Barbara Corpus (Du Bois et al. 2000-2005). The comparison shows a high level of agreement on the identification of continuative and conclusive boundaries.

Examples (1a-d) show how different segmentations lead to different possible speech acts. The ones chosen here are not the only possible ones, but segmentation is always the first step that restricts the performance and the perception of possible speech acts. It is easy to observe that segmentation determines also syntactic interpretations. Among other evident differences, while in (b) and (c) *People* functions as subject, this is not the case for (a) and (d), where the same word performs an autonomous speech act and therefore constitutes an utterance by itself.

Examples (2) and (3) below and their respective audios (extracted from Cavalcante *et al*. (2018))[2] present terminated sequences formed by two prosodic units. In example (2), the first unit is non-illocutionary and ends with a continuative prosodic boundary. The illocution is conveyed by the second unit ("you can only take it for eight weeks"), which carries the pragmatic and prosodic autonomy of the sequence. Conversely, in (3) the first unit is clearly illocutionary ("one of his brother's sons"), while the second is non-illocutionary and, for that reason, non-autonomous.

(2) afammn01[5]
*LYN: if you want to / you can only take it for eight weeks //

(3) afamdl01[115]
*RIC: one of his brother's sons / or something //

3.2 Utterances and stanzas

According to L-AcT, there are two different types of terminated sequences: utterances and stanzas. The main distinction between them is not directly related to the number of prosodic units that form the terminated sequence, but to the number of illocutionary patterns performed. Utterances are formed by a single pattern with one illocutionary unit (or a patterned illocution, as will be shown later), while stanzas have at least two juxtaposed patterns, each one with its own illocution, linked together by a continuation prosodic signal.

The following examples (4-6) present utterances (4) and stanzas (5 and 6), formed exclusively by their illocutionary units. In these examples, the utterance features a single prosodic unit, while stanzas carry two or more units. The examples are tagged according to L-AcT's annotation conventions; the illocutionary unit of the utterance and the last illocutionary of the stanza is the *Comment*

[2] All the examples can be listened in Cavalcante *et al.* (2018), that can be downloaded from www.c-oral-brasil.org > corpora > Minicorpus Inglês Americano (2018).

(COM), while the other illocutionary units of a stanza are *Bound Comments* (COB), to show that they partake of a unique terminated sequence.

(4) afamdl04[28-32]
*KAR: The Substitute Wife //=COM=
*JUL: oh //=COM= was that any good //=COM=
*KAR: it was very good //=COM=
*JUL: oh good //=COM=

(5) afamdl02[141]
*PAM: I look down on my body /=COB= and I fell like I'm in a spaceship //=COM=

(6) afamdl02[181]
*DAR: be a /=SCA= doctor /=COB= or a screen writer /=COB= or an actress /=COB= or a philanthropist /=COB= or an explorer //=COM=

Utterances and stanzas sub-patterns can also present a variable number of different types of non-illocutionary units, as in examples (7-10). These units are annotated with different tags that will be explained better below: *Topic* (TOP), *Parenthetical* (PAR), *Appendix of Topic* (APT), *Appendix of Comment* (APC) and *Auxiliary* (AUX), a general tag indicating that the unit has the function to regulate the communication, i.e. it is what in other frameworks are called Discourse Markers. Besides those information units, the tag (SCA), which means *scanned unit*, indicates an intonation unit which is part of a greater information unit, and (TMT), *Time Taking*, indicates a filled pause.

- Utterances carrying non-illocutionary units

(7) apubdl01[15]
*RAN: right now /=TOP= like you say /=PAR= maybe it is better for you to do some of that stuff //=COM=

(8) afamdl01[115]
*RIC: one of his brother's sons /=COM= or something //=APC=

- Stanzas carrying non-illocutionary units

(9) afamdl02[58]
*PAM: and then it was interesting /=COB= cause reading I did /=TOP= after that /=APT= substantiated that experience //=COM=

(10) afammn02[10]
*ALN: &he /=TMT= flew down to Mexico City /=COB= &he /=TMT= we &c [/1]=SCA= think of the name of my hotel /=COB= which wouldn't mean anything now /=PAR= but we ended up in a /=SCA= fabulous hotel /=COB= &he /=TMT= first night /=TOP= we were very unhappy with our rooms /=COB= we got down there /=COB= and the next morning /=TOP= Buddy /=TOP= who 's a /=SCA= early riser /=PAR= anyhow /=PAR= was probably up /=SCA= four o'clock /=COB= and he went down there complaining to the manager /=COM= so //=AUX=

These examples allow the understanding of a series of important questions regarding the composition of terminated sequences that can be used to explain important aspects of the organization of speech according to L-AcT: (a) Why does the speaker choose to produce an utterance (or a sequence of utterances) instead of a stanza and vice-versa? (b) What are the differences among non-illocutionary units? (c) How are the illocutionary and non-illocutionary units organized inside a terminated sequence?

A speaker produces an utterance in order to convey a single actional pattern towards another individual, and produces a stanza to convey a sequence of actional patterns inside the same terminated sequence. However, there is a more subtle distinction between the two possibilities: while the utterance places emphasis on the very illocution that it conveys and can be seen as an online program, stanzas are better understood as the expression of an ongoing textual elaboration, conveyed through a sequence of "weakened" and homogeneous illocutions, linked by a signal of continuity, and programmed one after the other. Stanzas normally take place in situations in which there is a lower level of interaction, and the speaker can focus on the semantic elaboration of the text. Therefore, stanzas tend to be more frequent in monological contexts (Cresti, 2005; Raso & Mittmann, 2012; Mello, 2014). For the aforementioned reasons, stanzas differ from a sequence of utterances, both from a prosodic and a cognitive point of view.

L-AcT recognizes the existence of two major types of non-illocutionary units: textual and dialogic units (or *Auxiliaries*, AUX). Textual units integrate the semantic text of the terminated sequence, AUX regulates the interaction among the

participants[3]. Each different information unit has its specific function, a specific prosodic form, and its own distributional behavior.

Example (7) starts with a *Topic* (TOP), the most frequent textual units. TOP is always positioned before the illocution and provides the cognitive domain for its interpretation, which is conveyed by a prosodic prominence, whose forms have been studied in several works (see at least Author et al., 2017 and Cavalcante *et al*., in press). Besides TOP and COM units, example (7) presents also a *Parenthetical* (PAR), a textual unit used by the speaker to make metalinguistic comments about the text (Tucci, 2010). Examples (8) and (9) show an *Appendix of Comment* (APC) and an *Appendix of Topic* (APT), respectively. The appendix units integrate the text of COM or TOP.

For our goals, there is a last unit that needs to be introduced: the *Multiple Comment* (CMM). CMMs are complex patternized illocutions, they are not juxtaposed as COBs are. CMMs are two (rarely more) illocutionary units that together build an illocutionary pattern with a holistic interpretation: reinforcement, comparison, different logic relations conveyed by prosody without a lexical operator, and functional calling are some examples of possible patterns. They can occur both in utterances and stanzas, and they represent more complex illocutionary solutions, both from a cognitive and a prosodic point of view.

Example (11) brings a sequence of reinforcement CMMs, which conveys the same illocution twice. It is interesting to notice that usually it is the second reinforcement CMM that has a more semantically elaborated text. Example (12) features a stanza with an adversative CMM illocutionary pattern. Example (13) consists of a stanza with three CMMs that convey a list illocutionary pattern.

It is worth noticing that in this stanza there is also another textual unit, the *Locutive Introducer* (INT). This unit signals that what follows must be interpreted as a whole (as in lists) or according to pragmatic coordinates that are different from the *hic et nunc*, as in reported speech.

(11) afammn01[2]
*LYN: no /=CMM= I don't have my equipment at all //=CMM=

[3] We will not differentiate the AUX subclasses for the purposes of this paper (see Raso & Vieira, 2016; Raso & Ferrari, 2020; Raso *et al*., 2022, for fine grained differentiations).

(12) afammn05[68]
*COR: things are meant to come out /=CMM= not go in /=CMM= in that hole //=COM=

(13) afamm02[25]
*ALN: we had /=INT= four or five of 'em /=COB= Aaron had one /=CMM= I had one /=CMM= Mike had one //=CMM=

Both in utterances and stanzas, there are no mandatory units other than the illocutionary ones. The speaker decides whether to include optional units based on her/his communicative needs. For instance, if the speaker wants to mark that the illocution conveyed by a given COM refers to a cognitive domain different from the one provided by the immediate context, she/he can include a TOP before it; if the speaker wants to make a comment on the main content of the utterance, he can insert a PAR. By doing so, the speaker creates an *informational pattern* – i.e., a set of information units formed by an illocutionary unit (the core of the pattern), around which the non-illocutionary units gravitate and to which they are cognitively and prosodically subordinated. This means that the speaker can organize, at a cognitive level, the information pattern she/he prefers, producing functional units conveyed by prosodic means. While an utterance presents only one informational pattern, each one of the illocutionary units of a stanza may have its own informational *sub-pattern,* i.e., a configuration formed by a COB and the non-illocutionary units specifically related to it.

The concept of information pattern (and sub-pattern) allows a better understanding of what happens behind the elaboration of a terminated sequence. This should not be seen just as a sequence of pragmatically independent units, which are related to one another only on syntactic or semantic levels. All textual units are meant to pragmatically and cognitively support the illocutionary one, since they are built as a specific program.

It is possible, therefore, to describe terminated sequences not only in terms of the number of their units, but also in terms of the number and types of information patterns they perform. Examples (9) and (10) show different sub-patterns. In example (9), the first COB (which does not form a sub-pattern with any non-illocutionary unit) is followed by a [TOP-APT-COM] sub-pattern. In example (10), there are five different sub-patterns: ([AUX-COB], [AUX-COB-PAR], [AUX-TOP-COB], [TOP-TOP-PAR-PAR-COB] and [COM-AUX]), while the third COB (“fabulous hotel /”) and the fifth one (“we got down there /”) do not form sub-patterns.

4. The methodology for the study of the information patterns

This section explains our methodology for the study of the information structure of spontaneous speech in individuals with schizophrenia. As has been pointed out, it was especially designed to overcome the fact that there is no specific control corpus for C-ORAL-ESQ. The patient data were therefore compared to those of the general reference corpus of Brazilian Speech Portuguese, the C-ORAL-BRASIL corpus.

The methodology does not aim to compare different speakers directly, but specific structures that can be found in the recordings of both C-ORAL-ESQ and C-ORAL-BRASIL. These specific structures are stanzas with the same number of illocutionary units. Our goal is to investigate the degree of complexity of these structures when they are produced in monologues – a context in which speech is less dependent from a specific situation and topic. Thus, stanzas with one COB in C-ORAL-ESQ are compared to stanzas with one COB in C-ORAL-BRASIL, stanzas with two COBs are compared to stanzas with two COBs, and so on. This strategy allows the comparison of stanzas with the same level of potential complexity, since they are formed by the same number of nuclear units, and then it is possible to analyze their degree of complexity, which can be measured by the richness of textual and dialogic units subordinated to the COB, as well as the number of CMMs. By doing so, it is viable to look for statistical differences in the use of specific information units and try to induce some cognitive and prosodic possible differences between individuals without mental disorders and individual with schizophrenia.

Costa (2020), which followed our methodology for the comparative study of stanzas, elaborated several Python scripts to extract results not only about information structure, which is the main goal of this work, but also about silent and filled pauses number and duration, lexical information and syllabic structure.

The first premise taken into consideration when we formulated our methodology is that different linguistic phenomena may be influenced by different factors. For instance, lexical variability of a corpus is in direct relation with the variability of its topics, while the illocutionary variation is due to the variation of speech contexts and degree of interactivity (Moneglia, 2011). In fact, in interviews regarding different topics we usually find a large lexical variation, but the speech acts performed by the speakers tend always to be the same (mainly questions, answers, assertions, and a few others). This is because speech context determines speaker's communicative needs and therefore elicit mainly a specific set of illocutions.

According to a series of studies carried on in the C-ORAL-BRASIL project and at the LABLITA lab, it has been found that a large set of properties related to information structure complexity (number of words, tone units, illocutionary and non-illocutionary units in a terminated sequence, among others) varies according to the degree of interaction between the participants (Cresti, 2005; Raso & Mittmann, 2012). If on the one hand, the more interactive the text is, the greater its illocutionary variability is; on the other hand, the less interactive the text is, the more reduced its illocutionary variation is and the more complex the information structure of the different patterns is. The variation induced by the degree of interaction can be appreciated in the following tables and graphics, adapted from the C-ORAL-BRASIL and C-ORAL-ROM statistics.

Table 1. Complexity of the dialogic turns of C-ORAL-BRASIL

	Terminated sequences / dialogic turn			Words / dialogic turn		
	Min	Max	General mean	Min	Max	General mean
Conversations	1.19	2.12	1.46	4.38	14.01	7.45
Dialogues	1.46	3.53	1.83	6.36	25.25	9.64
Monologues	1.89	90.00	3.01	12.85	44.94	28.56

Adapted from: Raso & Mittmann, 2012.

Table 1 shows the mean values of terminated sequences per dialogic turn and words per dialogic turns in C-ORAL-BRASIL texts. The first aspect to be noticed on the table is the great comparative difference in all measurements between conversations and dialogues, on one hand, and monologues, on the other hand. The numbers clearly indicate that, the less interactive a situation is, the greater the complexity of the dialogic turn in it is. It is worthwhile to notice that in C-ORAL-BRASIL we find a monologue with a mean of 90 terminated sequences per turn (which actually means it is formed by only one dialogic turn), while the general mean of this measure is 3.01. This discrepancy is due to the fact that, very often, the interlocutor constantly interacts with the person that is producing the monologue using small and structurally simple utterances which function as backchannels.

Table 2 portrays the mean number of tone units per terminated sequence, words by terminated sequence and words by tone units in C-ORAL-BRASIL. Dialogical interactions present smaller values on all three measurements in respect with monological interactions: 1.60 vs. 2.61 tone units per terminated sequence,

5.19 vs. 6.16 words per terminated sequence and 3.24 vs. 3.63 words per tone unit.

Table 2. Complexity of the terminated sequences of C-ORAL-BRASIL in familiar/private context

Context	Prosodic units / terminated sequence	Words / terminated sequence	Words / prosodic unit
Dialogical	1.6	5.19	3.24
Monological	2.61	9.48	3.63

Adapted from: Raso & Mittmann, 2012.

Table 3 presents the proportion between terminated sequences formed by a single prosodic unit and terminated sequences with at least two prosodic units in C-ORAL-BRASIL and in all four branches of C-ORAL-ROM. While terminated sequences with a single prosodic unit are necessarily utterances formed just by the illocutionary unit (*simple terminated sequence*, in the table), terminated sequences with two or more prosodic units (*complex terminated sequence*) can be both utterances with additional non-illocutionary units and stanzas with or without non-illocutionary units. The same tendency can be observed in all five corpora: clearly more than 50% of all terminated sequences in monologues are complex, while more than 50% of utterances of dialogical contexts are simple.

Table 3. Proportion between simple and complex terminated sequences in C-ORAL-BRASIL and C-ORAL-ROM

Context	Type of terminated sequence	Brazilian Portuguese	European Portuguese	Italian	Spanish	French
Dialogical	Simple	58.7%	50.2%	52.0%	57.8%	69.2%
	Complex	41.3%	49.8%	48.6%	42.2%	30.8%
Monological	Simple	43.2%	32.4%	30.5%	32.4%	44.1%
	Complex	56.8%	67.6%	69.5%	67.6%	55.9%

Adapted from: Raso & Mittmann, 2012, p. 191; Cresti, 2003, p.222.

It is worth mentioning that these numbers would show a much greater difference between dialogic and monologic texts, if they did not include dialogues with low degree of interaction, like interviews or chats, and if they excluded phenomena like backchannels in the monologues, that are very frequent and strongly increase the number of simple utterances.

It should be also considered that a significative part of complex terminated sequences is formed just by the illocution and one or more dialogic units, which are interactional information units that do not affect the textual complexity of the terminated sequence. Figure 1 explores this aspect by taking a deeper look at the information structure of *complex terminated sequences*. It distinguishes terminated sequences that present, besides the illocutionary unit, (i) only dialogic units, or (ii) textual units (plus eventual dialogical units). The figure shows that in monologues the number of complex terminated sequences with textual units largely overcomes the number of those ones with only dialogic units (71% vs. 29%). Conversely, in dialogic contexts, there are more complex terminated sequences with only dialogic units than those ones with textual units (58% vs. 42%). The presence of textual units indicates much more prosodic and cognitive complexity than the presence of dialogic units, which is an important consideration for our goals in this paper.

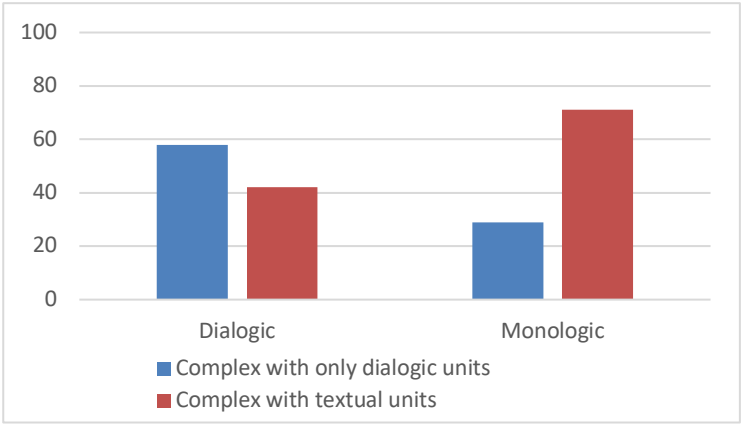

Figure 1. Proportion between complex terminated sequences with only dialogic units and complex terminated sequences with textual units in C-ORAL-BRASIL

Adapted from: Raso & Mittmann, 2012.

These data show that monologic speech strongly induces the complexity of terminated sequences not only on the textual level (*more words per terminated sequence* and *more words per intonation unit*), but also on the informational level (*more intonation units per terminated sequence*, *more complex utterances* than *simple utterances*; *more complex terminated sequences with textual units* than *complex terminated sequences with only dialogical units*).

This correlation, which is expressively consistent among different languages, can be easily explained by the fact that, the more actional an interaction is, the

more the participants are focused on the activity they are executing in the specific context, rather than on the linguistic structures they use to communicate with the interlocutor as a support for the activity. On the opposite end, the less actional an interaction is, the more focused on textual elaboration it is; this textual elaboration is largely independent from a specific situation the participants are in. This explains why all monologues tend not only to have stanzas, but also very complex patterns, regardless of their topic or the specific situation that leads a participant to talk more than the others.

We claim that this correlation validates our methodology to compare the speech produced in these kinds of structures from the two corpora. Since the complexity of an information pattern is mainly determined by the degree of interaction, the main factor to be controlled while choosing the data to analyze is precisely the degree of interaction between the participants, which must be low, and not the topic or the specific situations, which clearly differentiate C-ORAL-ESQ and C-ORAL-BRASIL. In fact, while C-ORAL-BRASIL tried to cover the greatest situational variability available, thus presenting a large variety of topics, C-ORAL-ESQ features a specific situation and topic.

That being said, there is another very important reason for choosing the stanza as the locus of investigation of the information structure in individuals with schizophrenia (and thus for the reliability of the methodology for comparison proposed here): the very nature of stanzas, which are terminated sequences in which the speaker is not focused on the accomplishment of a speech act, but rather on the elaboration of a text. As previously argued, stanzas convey sequences of illocutions of the same type, usually from the class of Representatives (that includes illocutionary types such as Assertion, Description, Explanation, Narration and Quotation; Cresti, 2020), especially useful to semantic/textual elaboration. The homogeneity of illocutions induced by monological contexts is a very desirable aspect in the comparison of information structures from different groups, because it reduces and normalizes the illocutionary variation, therefore eliminating another problem for the comparison of the information structure.

In our view, these arguments justify the comparability of textual complexity in stanzas, even if the structures are extracted from two non-comparable corpora. The main conclusion to be drawn from our arguments is that, provided that a stanza is produced in a monological context, the specific type of speech situation in which a stanza is produced does not play a fundamental role on the elaboration of its informational structure. This is the reason we argue that the structural and content differences between the C-ORAL-ESQ and C-ORAL-BRASIL corpora do not tame the results we will present.

Nevertheless, there is another crucial aspect to be controlled in order to study the information structure in stanzas, as we have briefly mentioned before: the number of illocutionary units of a stanza. As already pointed out, a stanza is formed by two or more illocutionary units, and each one can be integrated by one or more non-illocutionary units, creating a sub-pattern inside the stanza. Therefore, the complexity of a stanza is given not only by the number of illocutionary units, but mainly by the extension and complexity of each sub-pattern. Indeed, each illocutionary unit of a stanza can be seen as a nucleus of a potential sub-pattern, which can enrich the information structure of the stanza itself.

That is why it is not sufficient to compare different stanzas; we also need that they have the same number of illocutionary units. In other words, it is necessary to ensure that the stanzas to be compared have the same number of *potential sub-patterns*.

The analysis of the information sub-patterns of stanzas allows the appreciation of the cognitive capacity of the speaker to organize complex structures. Naturally, it would be possible to conduct this kind of analysis for the information patterns of utterances as well. However, as argued before, utterances are strongly dependent on the specific situation of the interaction, since their actional function is strictly related to the needs that the very moment of the interaction and the moves of the interlocutor require. We emphasize again that the more interactive the exchange is, the shorter the information patterns, the stronger the actional value and the more heterogeneous and unpredictable the illocution types. Conversely, stanzas, as typical monologic structures, constitute moments in which the speaker is relatively independent from the interaction and can elaborate her/his flow of thinking in a more textual and semantic way. She/he can elaborate a textual project and develop it for a larger time and without being much influenced by the interlocutor and the specific situation.

This is why, in order to compare the information structure of non-pathological and pathological speakers using non-comparable corpora we need to privilege the structure of the stanza. By doing so, we strongly neutralize the non-comparability of our corpora for phenomena relative to information structure.

An appropriate corpus of control should be built by interactions between physician and patient in a context of chronic disease that does not affect mental conditions or language (such as heart disease or diabetes), but allows familiarity between physician and patient, as a consequence of periodical consultations. Of course, the same diastratic characteristics of the patients should be preserved, which would mean, in a Brazilian context, that a comparable corpus should be built in a public health facility – the same social context of C-ORAL-ESQ. The compilation of a medic control corpus has proven to be a rather complex task,

marked by several bureaucratic barriers up to this point, and this has motivated us to work on the elaboration of a methodology that could be sound enough to allow comparative studies about prosody and information structure.

5. Statistical findings

This section briefly presents the most relevant findings in Costa (2022), which employed our methodology for the study of the information patterns of individuals with schizophrenia. In the next section we will propose a first analysis of Costa's work, that was limited only to the extraction of a set of linguistic measurements. Costa (2022) provides several measurements related not only to information structure, but which also included lexical metrics and disfluencies comparing the two corpora. Here we will focus only on information structure.

Costa studied all stanzas produced by 6 patients featured on recordings from C-ORAL-ESQ and then divided them in samples according to the number of illocutionary units they had (from 2 to 6). Then, the author selected 6 monologues from C-ORAL-BRASIL and randomly retrieved the same number of stanzas present in each sample of the C-ORAL-ESQ corpus. No variable due to diastratic characteristics of the speakers or monologue topic was taken into consideration. Table 4 shows the number of stanzas in each sample.

Table 4. Number of stanzas in C-ORAL-ESQ and C-ORAL-BRASIL samples

Illocutionary units	Number of stanzas in C-ORAL-ESQ samples	Number of stanzas in C-ORAL-BRASIL samples
2	113	113
3	46	46
4	26	26
5	6	6
6	10	10
Total	201	201

Adapted from: Costa, 2022: p. 50.

As shown by in Table 4, the most numerous – and thus most representative – samples are the ones containing 2 and 3 illocutionary units, respectively portraying 113 and 46 stanzas from each corpus. Indeed, not only the results observed in them are more consistent with one another, but it is here that most of the statistically significant results were found.

Figure 2 exhibits the sum of different informational patterns found in the samples. This graphic considers the whole samples, no matter the number of illocutionary units. As can be seen in Figure 2, there are more unique information patterns in C-ORAL-BRASIL samples than in C-ORAL-ESQ ones (124 vs. 96); Costa, however, does not mention whether this result has statistical relevance. Nevertheless, this result could indicate that the control group produces more complex stanzas than individuals with schizophrenia.

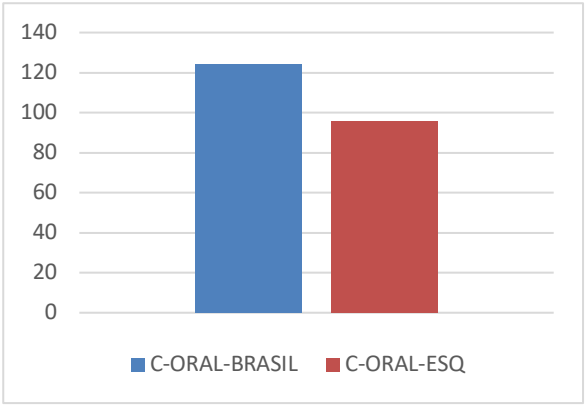

Figure 2. Number of different patterns in C-ORAL-BRASIL and C-ORAL-ESQ
Source: Costa, 2022, p. 106.

Figure 3 shows the most frequent information patterns in the stanzas of each corpus. The first three bar columns show that the most frequent stanza structures (COB-COM, COB-COB-COM, COB-COB-COB-COM) are the same for the compared samples. They, however, only portray illocutionary units and cannot elucidate much for our purposes. However, despite this overall similarity, the number of COB-COM patterns in C-ORAL-ESQ largely overcomes that in C-ORAL-BRASIL; this can be interpreted as a strong preference for structures combining only illocutionary units over structures that combine both illocutionary and non-illocutionary units and are therefore more complex. Another important issue to be observed is that all the other most frequent patterns from C-ORAL-BRASIL present textual units (INT, PAR and TOP), while those from C-ORAL-ESQ feature only dialogic units (AUX). As previously mentioned, textual units are much more complex than dialogic units since they perform more abstract and diverse functions, have larger lexical content and higher prosodic variability; therefore,

this result can be seen as another indication that individuals with schizophrenia tend to create less complex stanza sub-patterns than the control group.

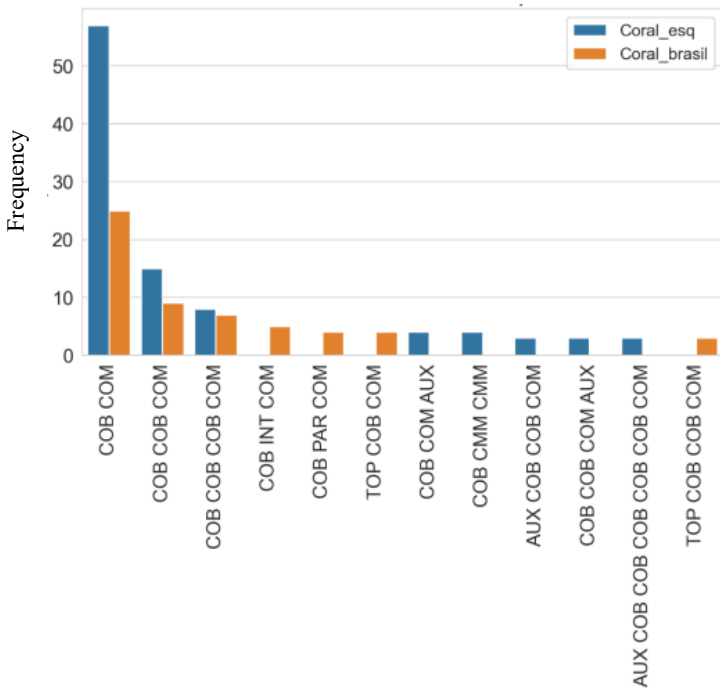

Figure 3. Most frequent informational patterns in C-ORAL-ESQ and C-ORAL-BRASIL
Source: Costa, 2022.

So far, we have shown overall data from the whole samples. Figures 4-9 compare structures in the two corpora according to their number of illocutionary units. Figure 4 shows the quantity of textual units on each sub-sample, i.e. comparing only structures with the same number of COBs.

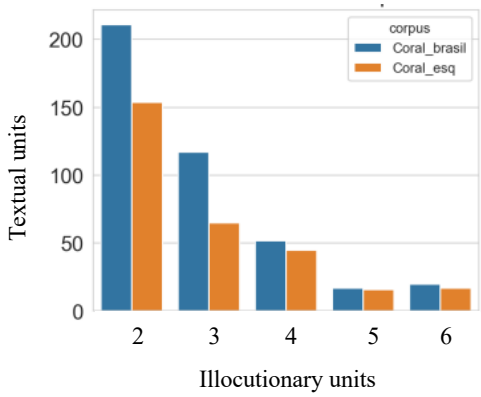

Figure 4. Total number of textual units in each sample of C-ORAL-ESQ and C-ORAL-BRASIL

Source: Costa, 2022.

As can be seen in the figure, all C-ORAL-ESQ sub-samples present fewer textual units, showing statistic relevance in the sample with 2 and 3 illocutionary units (which means 1 and 2 COBs) ($p = 0.0001$, $p = 0.001$, Mann-Whitney U Test).

Figures 5-9 show the distribution of TOP, PAR, INT, CMM and APC units in both corpora. C-ORAL-BRASIL brings more TOP, PAR, INT and CMM units than C-ORAL-ESQ in the first two samples, which are the largest and most representative ones. In the case of TOP units, the difference has statistic relevance for the first two samples ($p=1.64025e-05$ and $p=0.005$ in Mann-Whitney U Test). As for PAR and INT units, there is statistic relevance for the first sample only ($p=0,03$ for PAR unit and $p=0,01$ for INT in Mann-Whitney U Test). As for CMM units, there is statistic relevance for the second sample ($p=0,04$ in Mann-Whitney U Test). Finally, Figure 9 shows that C-ORAL-ESQ largely overcomes C-ORAL-BRASIL in the number of APC in the first two samples, although without reaching statistical relevance.

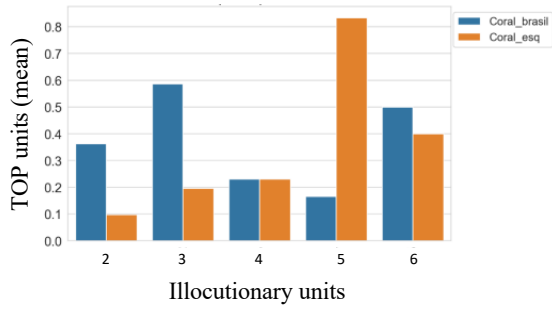

Figure 5. Mean of TOP units per illocutionary units
Source: Costa, 2022.

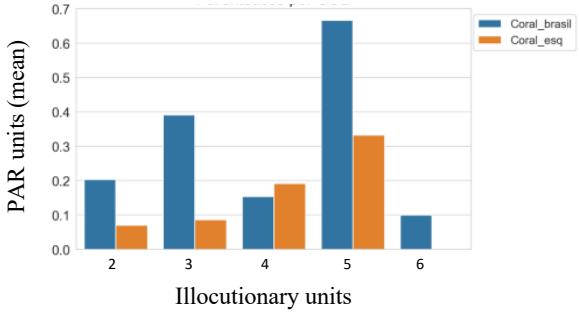

Figure 6. PAR units per COB
Source: Costa, 2022.

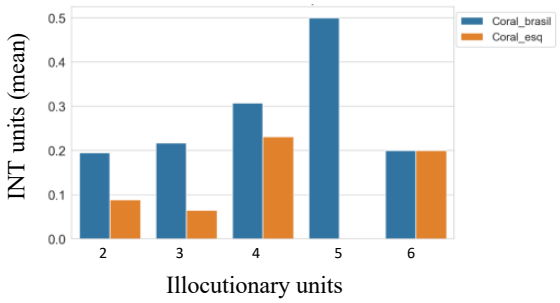

Figure 7. INT units per COB
Source: Costa, 2022.

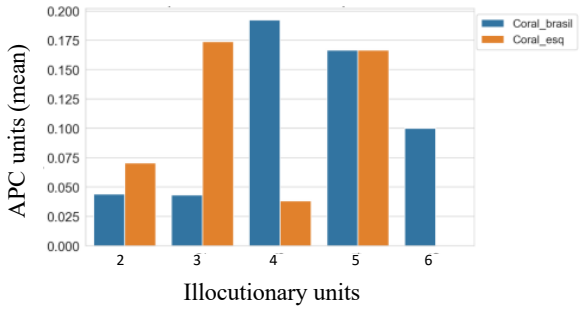

Figure 8. APC units per COB
Source: Costa, 2022.

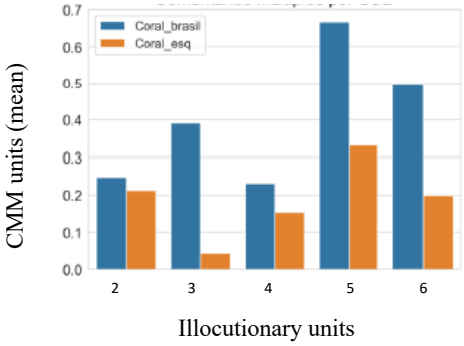

Figure 9. CMM units per COB
Source: Costa, 2022.

6. Statistical findings analysis

The previous section provides a first look at the basic properties of the information structure in stanzas produced by individuals with schizophrenia in comparison to those produced by the control group. More importantly, these results can be used to draw relevant insights about the cognitive and/or prosodic competence of both groups and could lead to further inquiries in this investigation. This section provides an analysis of the purely statistical findings by Costa's work.

It is important to notice that all data that were presented suggest that individuals with schizophrenia build stanzas with less complex information structures.

The simplification of the information structure occurs in at least three different levels[4].

- Stanzas with fewer non illocutionary units

The C-ORAL-ESQ sample presents a larger number of stanzas carrying only illocutionary units and a smaller number of stanzas combining both illocutionary and non-illocutionary units than the C-ORAL-BRASIL sample.

- Preference for dialogic units over textual units

The most frequent patterns of C-ORAL-ESQ that include non-illocutionary units present only dialogic units, which are much less complex than textual units, as previously discussed. On the other hand, the fact that some of the most frequent patterns of C-ORAL-ESQ include dialogic units could point to something related to the need to manage the relation with the interlocutor. However, this observation should be seen with caution, since Costa (2022) does not differentiate the existing types of dialogic units recognized by L-AcT, labelling all occurrences simply as AUX.

- Preference for APC over TOP

Special attention should be given to the more marked presence of TOP in the non-pathological speech sample and, on the other hand, the higher preference for APC in C-ORAL-ESQ. In fact, TOP and APC can be seen, both prosodically and cognitively, as opposing mirror units to some extent. TOP, as the unit that instantiate the cognitive domain for the illocution, is cognitively a very heavy unit. In order to perform a TOP, the speaker needs to evaluate its necessity and already have in mind the illocution for which the TOP must constitute the domain of reference. Besides, TOP has a very complex prosodic form. So far, three forms of TOP have been identified and modeled (Cavalcante *et al*., in press). All TOP forms feature

[4] Nervertheless, a fourth indirect observation, to be verified in further work, can be brought to attention. In their qualitative research based on the CIPPS corpus (Dovetto and Gemelli, 2012), Cresti et al. (2015) observe a noticeable reduction in the illocutionary variability of individuals with schizophrenia compared speech portrayed in the C-ORAL-ROM Italian. These observations are confirmed from one of the findings in our data, despite the fact that we chose to examine a structure that aims to avoid the effects of the illocutionary variation. In fact, in the C-ORAL-ESQ data we found a statistically significant reduction of CMMs compared to C-ORAL-BRASIL. Since CMMs are patterned illocutions that can be found both in utterances and stanzas, this reduced presence of CMMs may indirectly show a difficulty by schizophrenic speakers to vary the illocutionary value and also to perform illocutionary patterns, which for their very nature are more complex actional and prosodic structures.

a specific prosodic prominence that constitutes their nucleus and conveys the function of the whole unit. Therefore, TOP is a very complex information unit both from a cognitive and a prosodic point of view.

On the other hand, APC is a very simple unit. Firstly, it must always be cognitively given, while TOP can be new, and many times is used to detach the identification domain from the context (Hockett, 1958). Secondly, APC is always constituted by a short or very short sequence (many times just one word), that integrates the COM without adding a real semantic contribution, often repeating an information already given in the utterance or in a previous one. Thirdly, its prosodic profile is flat or falling, with low intensity.

Therefore, the fact that C-ORAL-ESQ features a much weaker presence of TOP (with a clear statistic difference) and a stronger APC presence (even if it is not statistically significant) can be a strong signal of cognitive and/or prosodic difficulties for the individuals with Schizophrenia. The quantitative results discussed here confirm what was already observed in previous qualitative works within the same framework (Cresti et al., 2015; Dovetto et al., 2015), which supports their validity.

7. Conclusion

In this paper we discussed a methodological design that allows the comparison of corpora that would otherwise, in principle, not be comparable. Our goal was to analyze the information structure in the speech of individuals with schizophrenia, taking the L-AcT framework as our theoretical orientation. This methodology is based on the choice of some specific structures, called stanzas, as comparable populations, since their complexity is independent, or at least much less dependent, from the distinguishing characteristics of the corpora analyzed.

The comparison undertaken has shown that schizophrenic speech is characterized by some challenges in the use of complex patterns and units that require more cognitive weight and prosodic complexity. There is a very common understanding that individuals with schizophrenia produce less melodic variation compared to individuals without mental disorders (Cohen *et al*., 2014), which is impressionistically measured by different psychometric scales, as SANS (Andreasen, 1989). However, many studies do not find strong and/or consistent results with statistical significance to describe the acoustic parameters that convey what is perceived as a "blunted speech affect" (Cohen *et al*., 2014; Alpert *et al*., 2020; Covington *et al*., 2012; Cohen *et al*.; 2013; Compton *et al*., 2018). On the other hand, one study seems to suggest an interesting direction for prosodic

investigation of schizophrenic speech: Martínez-Sanchez *et al*. (2015) observe that the melodic difference between the speech of individuals with and without schizophrenia reaches a clear statistical difference if measured not globally, but locally on intrasyllabic f0 movements. In fact, this study has found that individuals with schizophrenia produce a smaller percentage of prosodic peaks (measured by the percentage of syllables with tone changes ≥ 4ST), prosodic valleys (the percentage of syllables with tone changes ≤ -4ST) and with a diminished intrasyllabic trajectory (tone changes of the syllabic nuclei / duration) with respect to the control group. These data suggest that individuals with schizophrenia need more time to cover the same range movement of non-pathological speech, which can be related to what is presented in our study. In fact, both Topic and illocutionary units feature a clear prominence on the nuclear syllable(s). This opens the possibility for sound departing ground in further investigating the relation between prosody and information structure in schizophrenic speech.

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. doi: 10.1016/S0165-1781(00)00231-6

Andreasen, N. 1989. The Scale for the Assessment of Negative Symptoms (SANS): Conceptual and Theoretical Foundations. *British Journal of Psychiatry*, 155(S7): 49-52. doi: 10.1192/S0007125000291496

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

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 1999. *The Longman grammar of spoken and written English*. Harlow (Essex): Pearson Education Ltd.

Cavalcante, F.A., Raso, T. & Ramos, A.C. 2018. *American English Informationally Tagged Minicorpus*. http://c-oral-brasil.org/minicorpus_ingles_americano.html (accessed April 10, 2022).

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.

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. & Elvevåg, B. 2014. What do we really know about blunted vocal affect and alogia? A meta-analysis of objective assessments. *Schizophrenia Research*, 159: 533-538. doi: 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. doi: 10.1016/j.schres.2018.01.007.

Costa Jr., J. C. 2022. Padrão informacional de stanzas de pacientes com esquizofrenia. PhD thesis, Universidade Federal de Minas Gerais.

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. doi: 10.1016/j.schres.2012.10.005.

Cresti, E. 2000. *Corpus di Italiano parlato*. Firenze: Accademia della Crusca.

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. Amsterdam/Philadelphia: John Benjamins.

Cresti, E. 2009. La Stanza: un'unità di costruzione testuale del parlato. In *Atti del X Congresso della Società Internazionale di Linguistica e Filologia Italiana*.

Cresti, E. & Moneglia, M. (eds) 2005. *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam/Philadelphia: John Benjamins.

Cresti, E., Dovetto, F.M. & Rocha, B. 2015. Schizophrenia and prosody. First investigations. In C. Manfredi (ed), *Models and analysis of vocal emissions for biomedical applications - 9th International workshop*. Firenze: Firenze University Press, 139-142.

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 parlar matto. Schizofrenia tra fenomenologia e linguistica. Il corpus CIPPS*. Roma: Aracne.

Du Bois, J., Chafe, W., Meyer, C., Thompson, S., Englebretson, R. & Martey, N. 2000-2005. *Santa Barbara corpus of spoken American English*. Philadelphia: Linguistic Data Consortium.

Fleiss, J.L. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5).

Hockett, C.F. 1958. *A course in modern linguistics*. New York: Macmillan.

Izre'el, S., 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.

Izre'el, S., Mello, H., Panunzi, A. & Raso, T. 2020. Introduction: In search of a basic unit of spoken language. In: S. Izre'el, S., Author, A. Panunzi & Author. (eds), *In Search of Basic Units of Spoken Language. A corpus-driven approach*. Amsterdam/Philadelphia: John Benjamins, 1-32.

Jackendoff, R. 1977. *Syntax*. Cambridge (Mass): MIT Press.

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. doi: 10.1017/sjp.2015.85.

Mello, H. 2014. Methodological issues for spontaneous speech corpora compilation: the case of the C-ORAL-BRASIL. In T. Raso & H. Mello (eds), *Spoken corpora and linguistic studies*. Amsterdam/Philadelphia: John Benjamins, 29-68.

Mello, H., Raso, T., Mittmann, M.M., Vale, H.P. & 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: Editora UFMG, 125-176.

Moneglia, M. 2011. Spoken Corpora and Pragmatics. *Revista Brasileira de Linguística Aplicada*, 11(2): 479-519.

Moneglia, M. & Raso, T. 2014. Notes on Language into Act Theory (L-AcT). In T. Raso & H. Mello (eds), *Spoken Corpora and Linguistic Studies*, Amsterdam/Philadelphia: John Benjamins, 468-495.

Moneglia, M., Fabbri, M., Quazza, S., Panizza, A., Danieli, M., Garrido, J.M. & Swerts, M. 2005. Evaluation of consensus on the annotation of terminal and non-terminal prosodic breaks in the C-ORAL-ROM Corpus. In E. Cresti & M. Moneglia (eds), *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam/Philadelphia: John Benjamins, 257-274.

Panunzi, A., Gregori, L. & Rocha, B. 2020. Comparing annotations for the prosodic segmentation of spontaneous speech. In S. Izre'el, Author. Panunzi & Author (eds), *In search of basic units of spoken language: a corpus-driven approach* (Studies in Corpus Linguistics 94.). Amsterdam /Philadelphia: John Benjamins, 403-431.

Raso, T. & Mello, H. (eds) 2012. *C-ORAL-BRASIL I: Corpus de referência do Português Brasileiro falado informal*. Belo Horizonte: Editora UFMG.

Raso, T. & Mittmann, M.M. 2012. As principais medidas da fala. In T. Raso & H. Mello (eds), *C-ORAL-BRASIL I: Corpus de referência do Português Brasileiro falado informal*. Belo Horizonte: Editora UFMG, 177-221.

Raso, T. & Vieira, M. 2016. A description of Dialogic Units/Discourse Markers in spontaneous speech corpora based on phonetic parameters. *Chimera: Romance Corpora and Linguistic Studies*, 3: 221-249. Available at: <https://revistas.uam.es/index.php/chimera/issue/archive>.

Raso, T. & Ferrari, L.A. 2020. Uso dei Segnali Discorsivi in corpora di parlato spontaneo italiano e brasiliano. In R. Ferroni & M. Birello (eds), *La competenza discorsiva e interazionale A lezione di lingua straniera*. Canterano: Aracne editrice, 61-107.

Raso, T., Rilliard, A. & Santos, S. 2022. Modeling the prosodic forms of Discourse Markers. *Domínios de Linguagem*, 1436-1488.

Rocha, B., Raso, T., Ferrari, L., Mantovani, L. & Salgado, J. 2020. A corpus of Brazilian Portuguese speech by schizophrenic patients: preliminary observations. In Dovetto, F. (ed), *I sistemi instabili*. Canterrano: Aracne, 307-334.

Raso, T., Cavalcante, F., Mittmann, M. 2017. Prosodic forms of the Topic information unit in a cross-linguistic perspective: a first survey. In A. de Meo & F.M. Dovetto (eds), *Proceedings of the SLI-GSCP International Conference, 13-15 June, 2016*. Roma: Aracne.

Tucci, I. "Obiter dictum". 2010. La funzione informativa delle unità parentetiche. In: M. Pettorino, A. Giannini & F.M. Dovetto (eds), *La comunicazione parlata, 3, 2009, Napoli. Atti del Convegno Internazionale GSCP, 1.* Napoli: Università degli Studi di Napoli l'Orientale, 635-654.