Segmentation of the Speech Flow for the Evaluation of Spontaneous Productions in Pathologies Affecting the Language Capacity.
4 Case Studies of Schizophrenia

Valentina Saccone, Simona Trillocco
University of Florence

valentina.saccone@unifi.ti, simona.trillocco@unifi.ti

Abstract

This paper aims to present a multi-level analysis of spoken language, which is carried out through Praat software for the analysis of speech in its prosodic aspects. The main object of analysis is the pathological speech of schizophrenic patients with a focus on pausing and its information structure. Spoken data (audio recordings in clinical settings; 4 case studies from CIPPS corpus) has been processed to create an implementable layer grid. The grid is an incremental annotation with layers dedicated to silent/sounding detection; orthographic transcription with the annotation of different vocal phenomena; Utterance segmentation; Information Units segmentation. The theoretical framework we are dealing with is the Language into Act Theory and its pragmatic and empirical studies on spontaneous spoken language. The core of the analysis is the study of pauses (signaled in the silent/sounding tier) starting from their automatic detection, then manually validated, and their classification based on duration and position inter/intra Turn and Utterance. In this respect, an interesting point arises: beyond the expected result of longer pauses in pathological schizophrenic than non-pathological, aside from the type of pause, analysis shows that pauses after Utterances are specific to pathological speech when >500 ms.

Keywords: spontaneous speech, segmentation, schizophrenic speech

1. Introduction

Our main purpose is to broaden the pragmatic knowledge of pathological speech, starting from the observation of the prosodic profile of 4 patients with schizophrenia and highlighting their atypia in contrast to non-pathological speech. This work focuses on a functional and structural method, creating a standard to analyze and describe pathological spontaneous spoken language. For this reason, we present the elaboration of a visual structure of segmentation with Praat software (Boersma and Weenink, 2021), on which it is possible to develop different and parallel linguistic levels of analysis for the study of spoken language. We organized a layer grid, starting from an early distinction between sounding and silent, firstly using an automatic script and then manually validating the resulted segmentation. Further annotations have been added at different levels of analysis that are interconnected and allowed cross-layer observation of spoken data.

In more detail, we report here 4 case studies on schizophrenic patients based on the analysis of the Italian CIPPS corpus (Dovetto and Gemelli, 2013) compared to non-pathological spoken language of the Italian section of C-ORAL-ROM corpus (Moneglia, 2005).

Our theoretical framework is L-AcT, Language into Act Theory for the information structure of speech (Cresti, 2000; Cresti and Moneglia, 2010; Moneglia and Raso, 2014) in the reference point of pragmatics. In this perspective, the speech is naturally divided into linguistic units − easily identified by perception − that are called *utterances*. The Utterance[1] carries the meaning expressed by the speaker, is autonomous and independent (Cresti, 2000): it must necessarily have an illocutionary force (Moneglia and Raso, 2014). Its identification as the linguistic counterpart of a speech act (Austin, 1962; Cresti, 2000) allows us to make important observations on the prosodic characteristics of pathological patients and their strategies for the information articulation of speech.

The analysis of Information Units in schizophrenic speech is a continuation of works of Dovetto, Cresti and Rocha (2015) and Cresti and Moneglia (2017).

2. CIPPS Corpus

The CIPPS Corpus (Corpus of Italian Spoken Pathological/Schizophrenic)[2] is a collection of psychiatric interviews with 4 schizophrenic patients, anonymized with the letters A, B, C, and D. They experience different stages of the disease: A is in a pre-delusional condition of *Wahnstimmung* without hallucination; B suffers from paranoid schizophrenia with unstructured delirium without hallucinations; C has paranoid schizophrenia with structured delirium and hallucinations; D corresponds to the diagnosis of paranoid schizophrenia with delirium.

The corpus currently consists of 17 hours of recordings in an ordinary environment without any elicitation: three medical sessions for A (150 min), four sessions for B (238 min), two sessions for B (128 min), and one session for D (28 min). The four subjects are all males from Naples, with an age ranging from 35 to 45, and report standard Italian speech with some dialect inflections (more consistently pronounced in D). The corpus is currently being updated with new acquisitions of drug-resistant schizophrenic speech, in collaboration with the AOU of Naples Federico II[3]. The project is edited and coordinated by Dovetto and dedicated to "Non-standard Dialogic Speech Corpora" which also includes an innovative PhD scholarship (Dovetto et al., 2021).

---

[1] In this work, we use Utterance with capital letter because it is considered a unit of measurement for speech and segmentation on PRAAT. The same applies for Information Units.

[2] The CIPPS corpus results from the collaboration between the Scuola Sperimentale per la Formazione alla Psicoterapia (ASL NA1) and CIRASS - Centro Interdipartimentale di Ricerca per l'Analisi e la Sintesi dei Segnali of the University of Naples "Federico II".

[3] The contact person for the AOU is Prof. De Bartolomeis.

The recording sessions are in the form of dialogues/medical interviews between the patients and their doctor, and mainly consist of monological excerpts due to the low presence of the doctor's turns.

The recordings were manually transcribed (the transcription is available for the first 10 hours) with orthographic criteria based on Savy (2007), then implemented in Dovetto and Gemelli (2015) reporting different types of phenomena: vocal non-verbal phenomena, such as laugh, tough, breath, inspiration, tongue click and throat clearing; vocal non-lexical phenomena, as vowel or consonant lengthening, vocalizations and nasalization; and empty pauses, initially divided into small <sp> and large <lp>, then more finely classified in relation to specific thresholds.

3. Methods

The first step of spoken data processing has been analyzing the recording sessions through WinPitch software (Martin, 2004) for the text-sound alignment, and then through Praat software, to identify Utterances and, within them, Information Units and their exact prosodic boundaries.

More specifically, through Praat TextGrids, the audio files have been processed with a multi-level analysis, obtaining an incremental annotation with one information per tier:

- silent/sounding detection;
- Utterance identification with orthographic transcription, with the annotation of different vocal phenomena;
- Information Units identification;
- Tag of Information Units.

This annotation can be implemented with other and potentially unlimited levels, from phonetic and phonological phenomena (vowel lengthening, different types of vocalizations or nasalizations) to paralinguistic annotations (breathing/empty silences; tongue-clicks; cough, laugh, throat clearing, etc.).

After annotating, we differentiated between spontaneous speaking and other peculiar parts[4] of the clinical sessions, such as reading (for patients C and D) or drawing (for patient C).

3.1 Silent detection

The first tier is named "silences" in addition to a code that includes the letter identification of the patient and the number of the recording session[5]. The tier reports a distinction between sounding and silent stretch of the recordings. The method for the data processing is divided into three steps:

- noise removal, if necessary;
- automatic segmentation of the recording sessions in sounding and silent segments, with a preparatory adjustment of the dedicated Praat script based on the minimum intensity value per speaker;
- manual control of the automatic procedure by two evaluators.

To generate denoised audio files, we adopted a Praat tool that automatically elaborates a noise profile on the base of a selected time range inside the recording, and operates a spectral subtraction[6].

Regarding the segmentation in sounding and silent segments, the clearer the sound, the more the automatic procedure is reliable. However, due to the type of recording, the manual operation was still pervasive[7]. To assess the reliability, reproducibility and consistency of the segmentation, we carried out an agreement test between annotators, resulting in a rate of 0.85[8].

After the detection of pauses, we individuated a minimal threshold for silence in the value of 150 ms[9], according to the average duration of stop consonants (cf. Giannini, 2008; Dovetto and Gemelli 2013), and then we operated two different classifications: one on duration and the other on position criteria.

Concerning the duration, we operated a preliminary analysis based on the literature (references listed below in this paragraph) to identify significant thresholds; then, we considered equidistant thresholds in order to observe objective differences in the distribution of pauses. For this reason, we created two groupings related to intervals with different thresholds. In the first, the thresholds are distributed according to the following non-regular distances: 200 ms (Lea and Kloker, 1975; Duez, 1985); 250 ms, (Moneglia, 2005; Dovetto et al., 2021); 500 ms (Dovetto et al., 2021); 1000 ms and 5000 ms (Dovetto and Gemelli, 2013). Lastly, we added a 20000 ms threshold, which allowed us to identify very long pauses in patients with schizophrenia. In the second, the grouping follows regular intervals of duration to evaluate the distribution trend of pauses.

Regarding their position, and according to the literature, pauses have been tagged considering inter/intra Turns placement, and a further classification was adopted for the silences within the same turn (cf. *inter-tours* and *intra-tours* in Dodane and Hirsch, 2018; gaps and pauses in Heldner and Edlund, 2010; Fors, 2011). The resulting typology distinguishes between:

- T-pauses: inter Turns pauses;
- UT-pauses: inter Utterances pauses (inside the same Turn);

---

[4] In the medical interviews, patients (except for patient B) sometimes read texts previously written at home and discuss them with the doctor. In two cases (C and D), there is a description of a drawing; in one case (C), there is a dermatological examination describing a physical state, showing body parts to the doctor.

[5] The code consists of PZ, an abbreviation for "patient", plus an identification letter for each of the 4 patients (A, B, C, and D) and the number of the recording. The second recording of patient D, for example, is indicated by the PZD2 code.

[6] The method of spectral subtraction was defined in Boll (1979). The variant implemented in Praat is modeled after a script by Ton Wempe.

[7] Only the 11% of boundaries of the automatic detection remains unaffected.

[8] The test agreement has been made on a sample of PZD. On the base of the silent/sounding detection, we observed the manually verified boundaries comparing starting (t-min) and ending (t-max) times of silences. We adopted a fluctuation range of 150 ms, based on the minimum chosen threshold.

[9] For what concerns pauses under 150 ms, they are unlikely to seem relevant in monologues. Note that in Duez (1982) pauses were considered significant within the speech flow when <180 ms. The same threshold has been selected by CMU Open Source Speech in speech analysis (https://cmusphinx.github.io/).

- IU-pauses: intra Utterances pauses (between Information Units inside the same Utterance).

The observation of the two scales and mainly their interaction reveal important details about the behavior of pauses in schizophrenics. Among the various possible developments of the pauses analysis, there is the differentiation between empty silences and silences with paralinguistic annotations (cf. *respiratoires* and *non respiratoires* pauses in Fauth and Trouvain, 2018).

3.2 Utterance identification

The second tier is labeled "utterances" with the code of the patient. It reports the orthographic transcription of the speech flow with the annotation of specific vocal phenomena; the speech is here segmented into Utterances according to L-AcT.

Based on perception, it is possible to identify terminal breaks inside the speech flow that function as boundaries of interpretable units of the language. The theoretical framework we are dealing with has its core in the correspondence between pragmatic and prosodic units in speech, based on the empirical observation of linguistic corpora and tonal contour analysis (Cresti and Moneglia, 2010).

Each Utterance is filled with its transcription. Thanks to this, we can observe the presence or the absence of specific linguistic characteristics such as disfluency or retracting phenomena, and verify their percentage in schizophrenic speech Utterances. We can also calculate the number of Utterances and their length in terms of word numbers. Furthermore, it is possible to measure the stretch of speech[10] of each patient by correlating this tier to the sounding/silent value of the first one.

The orthographic transcription of the Utterances is internally segmented into Information Units, separated with non-terminal breaks.

3.3 Information Units identification

The third tier is labeled "words" with the code of the patient; it is used to segment the transcription of each Utterance in the corresponding Information Units (Moneglia and Raso, 2014)[11].

Inside Utterances, non-terminal boundaries show the information structure of speech underlining different strategies of language architecture. With the support of the prosodic configuration, we segmented the Utterances in non-autonomous units, i.e. the Information Units. The most significant clue to validate this phase is the pitch contour, both analyzed with Praat and WinPitch. Non-terminal boundaries can occur not only in the presence of pauses, but also concurrently with an f0 reset, intensity variations, or the change of the voice quality.

Below the "words" tier, a fourth tier indicates the tag of the Information Units according to L-AcT. It is named "info.units" together with the code of the patient. This layer of annotation allows us to link a word (or a series of words) to its pragmatic function, and easily identify the more recurring types of Informative Units used by speakers. This analysis permits the elaboration of precise statistics for schizophrenic speech, also and above all in comparison with non-pathological speech.

Figure 1 shows an example of the multi-level annotation described so far.
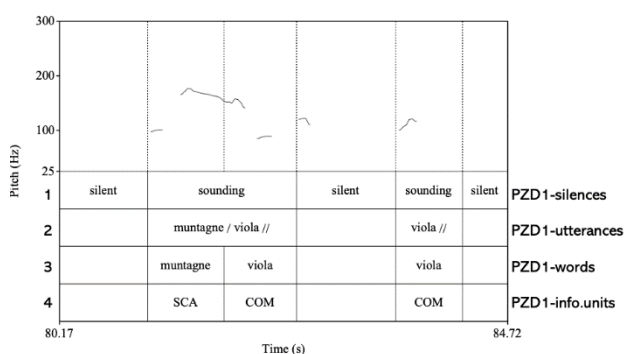

Figure 1. Example of the multi-level annotation in PZD1

4. Analysis

Based on the 3 levels of segmentation it is possible to extract data and information about: silent analysis; information structure of the language; comparison with non-pathological data and its measurements.

Thanks to the transcription, it is also possible to obtain measurements about lexical density, part of speech analysis (automatic PoS tagging), verbal/non-verbal utterances; disfluencies such as retracting phenomena.

We will discuss here in-depth data and results of silent analysis of CIPPS and its comparison with non-

---

[10] The stretch of speech includes silent/sounding only intra-turns, net of T pauses (see 4.1).

[11] As above (see 1.), the Utterance must necessarily have an illocutionary unit (= unit of Comment), the only one that can be interpreted as such in isolation. If the illocutionary unit is not accompanied by other elements, it is called *simple utterance*, otherwise *compound utterance*. The Information Units can be *textual* or *dialogic*. The first ones, of which the illocutionary unit is also part, constitute the semantic part of the Utterance, while the second ones (AUX = dialogic auxiliary) do not participate in the construction of the meaning of an Utterance but perform functions for its pragmatic success.

The textual units, in addition to the Comment, are the Topic (the identification domain of the Comment, and generally identified by three specific prosodic profiles, of which the most common in Italian presents ascending contour on the tonic and descending on the post-tonic), the Appendix of Comment and the Appendix of Topic (additions, often negligible, with descending or flat

prosodic profile and low intensity), the Parenthesis (a secondary textual level, generally with a lower intensity and higher rate) and the Locutive Introducer (which introduces meta-illocutions, the most frequent of which is the reported speech).

The dialogic units, on the other hand, are the Incipit (with short duration, high intensity, and variable profile, marks a contrast with the previous Utterance or a turn taking), the Conative (with short duration, high intensity, and descending profile, has the function of pushing the listener to take part in the dialogues, or stopping his uncollaborative behavior), the Allocutive (with short duration, low intensity and descending profile, identifies the interlocutor and establishes social cohesion), the Phatic (with short duration, low intensity, variable profile, keeps the communication channel open), the Expressive (prosodically variable, represents emotional support for the illocution) and the Discursive Connector (long duration, medium intensity, and variable profile, establishes a bond without contrast between two statements or subpatterns).

pathological data. All the measurements have been calculated on the patients, excluding the doctor's speech.

4.1 Silent analysis

Pause duration and collocation inside/between turns of conversation have been analyzed, thanks to the interaction between the different layers of segmentation. Pauses have been marked in a dedicated layer as described in 3.1 per type and divided into groups based on their durations.

First of all, it is interesting to notice the relation between the duration of pauses (silences) and the duration of the stretch of speech (silences+soundings inside the turns) per patient.

	Pauses	Stretch of speech	P/SoS
A	1582.3 ms	2669.2 ms	59.3%
B	3158.3 ms	13205.1 ms	23.9%
C	1381.9 ms	4810.1 ms	28.7%
D	235.8 ms	908.4 ms	25.9%

Table 1: Pause/Stretch of Speech

Data show that A's behavior stands out from the other patients and reflects his effort in communicating and keeping the turn (almost 60% "filled" with silences). This measurement increases its importance when evaluated in comparison with other data (such as non-pathological data) because it eliminates the T-pause influence on data, that is the most affected by the context in which the communication takes place. Even if the percentage of silence is exaggeratedly high just for A, there is a stronger presence of pauses also in the other patients than in non-pathological speech (Goldman-Eisler, 1961; Banfi, 1999; Heldner and Edlund, 2010).

Inside the turn, measurements of the four patients show a different trend for the two types IU and UT:

- IU-pauses follow 15.1% of IU and are mostly <1000 ms;
- UT-pauses follow 41.6% of UT and show a relevant peak of occurrence in the duration of 500-1000 ms.

A unique behavior is observed in patient B concerning UT-pauses: in his case, the percentage of UT followed by a pause raises to 61.1%, strongly influencing the mean percentage (35.1% without D's measurements) and prolonging his time of building turns.

For what concerns T-pauses, they follow 73.9% of T and are mostly <1000 ms. In this case, the behavior of patient

B strongly influences the mean percentage, because 70.9% of his silence is made only by T-pauses >1000 ms.

Moreover, our analysis shows that talking about pauses >1000 ms for CIPPS is likely to be reductive; above this threshold, we find pauses with duration >5 s or even >20 s. For a general overview of the frequency of pauses Table 2[12] shows the absolute number and the percentage (in brackets):

	IU Pauses	UT Pauses	T Pauses
A	312 (39%)	217 (27%)	273 (34%)
B	2289 (44%)	2656 (50%)	334 (6%)
C	739 (42%)	574 (32%)	455 (26%)
D	159 (40%)	113 (28%)	126 (32%)

Table 2: Frequency of pauses

4.2 Information structure analysis

Ongoing analysis shows that the four patients' speech has a clear attitude for simple Utterances; in fact, nearly 50.7% of the CIPPS utterances are filled by a single Information Unit. More precisely, the percentage differences between the four patients are minimal: 54.3% for A, 57.7% for B, 45.2% for C, and 47.9% for D. Even if in two cases (C and D) the speakers produce more compound utterances, their number is still low.

This means that the schizophrenic internal structure of the Utterance is usually poor, and the autonomous illocutions are mainly not accompanied by other textual or dialogic units, as in the following example (where the double slash // indicates the terminal boundary, i.e. the perceivable end of the utterance):

(1)   PZA1: questa è la domanda //

[*this is the question//*]

Further analysis will show new characteristics of the schizophrenic speech concerning the Information structure after completing the annotation of the units following L-AcT (Moneglia and Raso, 2014).

4.3 Comparison with non-pathological data

CIPPs data have been compared with non-pathological spoken data collected through previous linguistic analysis of spontaneous speech, namely on the Italian section of C-ORAL-ROM corpus within L-AcT theoretical framework, selecting a subset of male speakers[13].

---

[12] Preparatory statistical analysis (Kruskal-Wallis and Dunn tests) regarding pause duration per type (IU, UT, T) highlight the lack of homogeneity in the corpus; the only non-significant difference appears in A, C, and D measurements of IU-pauses. The analysis has been carried out by Lorenzo Gregori.

[13] The subset was chosen in particular to have gender homogeneity with the schizophrenic corpus, where the subjects are all males.

As already mentioned, the doctor-patient relationship conditions the speech properties. More specifically, one of the main differences between medical interviews and spontaneous dialogues is the turn-taking rate. In spontaneous spoken language, the hearer tends to answer before the very end of another speaker's turn, as soon as he/she understands the interlocutor's intent. Non-pathological conversation often appears to be characterized by overlaps, while in psychiatric sessions the doctor limits himself to a few backchannels and lets the patient speak.

To avoid this asymmetry, the distinction between different types of pauses results relevant and permits to compare only the two sets of IU- and UT-pauses of pathological and non-pathological.

The two plots below (Figures 2 and 3) compare schizophrenic and non-pathological speech concerning the types of pauses divided by their duration.
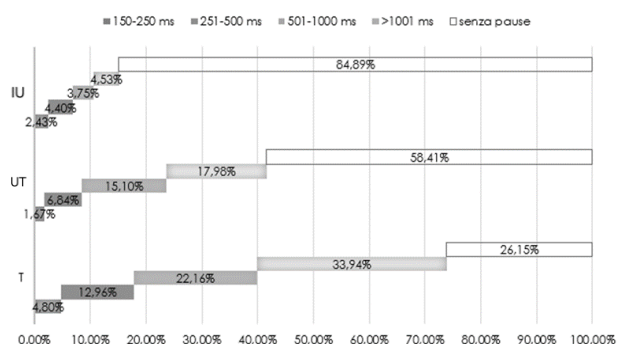

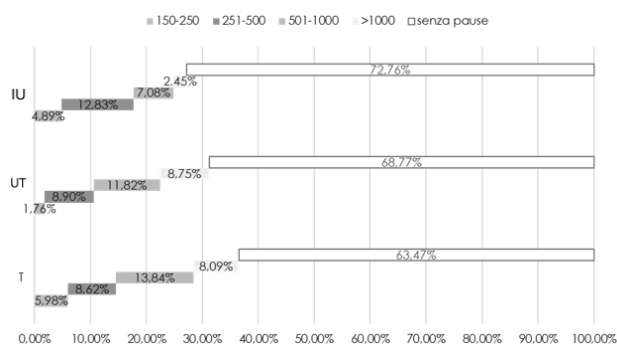Figure 2. CIPPS mean duration of pauses (frequencies)

Figure 3. non-pathological mean duration of pauses (frequencies)

Regarding the Information structure analysis, the presence of simple and compound utterances in non-pathological speech reveal interesting observations. In fact, the percentage of complex utterances in non-pathological speech is nearly 68% (see also Cresti and Moneglia, 2005), that is greater than in the four patients.

5. First results

Even if we are dealing with 4 case studies and the research is ongoing, our analysis already revealed interesting and coherent pieces of information on schizophrenic speech, which immediately suggest the characteristic atypia of this type of speech. In fact, the trend of pauses in CIPPS is clearly perceived as different from the non-pathological speech and, despite the non-homogeneity in data collection,

highlights a particular mental organization about the *position*, and therefore the *function*, of pauses within the Utterance.

Expected result, consistent with the literature (among the others, see Banfi, 1999), is that pauses of pathological schizophrenic speech are generically longer than non-pathological aside from the type (IU, UT or T).

A distinction between IU- and UT-pauses can be stated: IU-pauses match the non-pathological trend for what concerns their durations; UT-pauses >500 ms are rather more numerous than the non-pathological pauses. More in detail, for the control group: with increasing duration i. the incidence of IU pauses significantly decreases (from 70% with 937 pauses in the range 250-500 ms to 36% with 214 pauses in the range 1000-5000 ms); ii. the UT pauses increase (from 22% to 55% in the two considered ranges). Instead, the trend for schizophrenic subjects is different: i. IU pauses are quite more than UT pauses for the duration between 250 and 500 ms (53% with 1055 IU pauses vs 35% with 700 UT pauses); ii. IU pauses significantly decrease in the range 500-1000 ms with a clear preponderance of UT pauses (57% of UT pauses vs 30% of IU pauses i.e. 1450 vs 779 occurrences), similarly to the trend between 1000 and 5000 ms.

This means that a greater presence of pauses inside the Stretch of Speech in CIPPS underlines the difficulty of these patients in speech processing. The silence is a symptom not only of lexical retrieval (Dovetto and Gemelli, 2013), but also of a weak Information structure.

Finally, we remark that all the observations are made thanks to the visual structure of Praat. In fact, the cross layers interaction allows an in-depth analysis of schizophrenic speech, and it is easily implemented according to the linguistic aspect of interest (lexical, morphological, etc).

6. Bibliographical References

Banfi, E. (1999). Pause, interruzioni, silenzi. Un percorso interdisciplinare. In *Labirinti*, 36, Trento, Dipartimento di Scienze Filologiche e Storiche.

Boersma, P. and Weenink, D. (2021). Praat: doing phonetics by computer [Computer program]. Version 6.1.50, retrieved 20 June 2021 from http://www.praat.org/

Boll, S.F. (1979). Suppression of acoustic noise in speech using spectral subtraction. In *IEEE Transactions on ASSP* 27:113-120.

CMUSphinx, Open source speech recognition toolkit, https://cmusphinx.github.io/

Cresti, E. (2000). Corpus di italiano parlato. In *Studi di grammatica italiana pubblicati dall'Accademia della Crusca*. Firenze: presso l'Accademia della Crusca.

Cresti, E. and Moneglia, M. (2010). Informational Patterning Theory and the Corpus based description of Spoken language. The compositionality issue in the Topic Comment pattern. In M. Moneglia. and A. Panunzi (Eds), *Bootstrapping Information from Corpora in a Cross Linguistic Perspective*, Firenze: Firenze University Press, 13-46.

Cresti, E. and Moneglia, M. (2017). Prosodic Monotony and Schizophrenia. In *Lingua e patologia*. Napoli, Aracne:147-197.

Dodane, C. & Hirsch, F. (2018). L'organisation spatiale et temporelle de la pause en parole et en discours. In *Langages*, 211:5-12.

Dovetto, F.M. and Gemelli, M. (2013). Il parlar matto. Schizofrenia tra fenomenologia e linguistica. Il corpus CIPPS. Napoli, Aracne.

Dovetto, F.M., Cresti, E., and Rocha, B. (2015). Schizofrenia tra prosodia e lessico. Prime analisi. In *Studi Italiani di Linguistica Teorica e Applicata*, XLIV. Pacini Editore:486-507.

Dovetto, F.M., Guida, A., Pagliaro, A. C., Guarasci, R., Trillocco, S., Sorrentino, A. and Raggio, L., (forthcoming). Corpora di italiano parlato patologico dell'età adulta e senile: CIPPS, CIPP-ma, CIPP-mci. In *Congresso Internazionale della Società di Linguistica Italiana*, LIV, online, 8-10 settembre 2021.

Dovetto, F.M., Guida, A., Pagliaro, A. C. and Guarasci, R., (forthcoming). Silences and disfluencies in a corpus of patients with Alzheimer's Disease (CIPP-ma). In R. L. Rose and R. Eklund, *Proceedings of DISS 2021*, Saint Denis, Université Paris VIII Vincennes:121-124.

Duez, D. (1982). Silent and non-silent pauses in three speech styles. In *Language and Speech* 25(1):11-28.

Duez, D. (1985). Perception of Silent Pauses in Continuous Speech. In *Language and Speech*, 28:377-389.

Fauth, C. and Trouvain, J. (2018). Détails phonétiques dans la réalisation des pauses en français: étude de parole lue en langue maternelle vs en langue étrangère. In *Langages*, 211:81-95.

Fors, K.L. (2011). Pause length variations within and between speakers over time. In *Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue*, Los Angeles:198-199.

Giannini, A. (2008). I silenzi del telegiornale. In M. Pettorino, A. Giannini, M. Vallone and R. Savy (eds), *La comunicazione parlata (I)*, *Atti del Congresso Internazionale (Napoli 23-25 febbraio 2006)*. Napoli, Liguori:97-108.

Goldman-Eisler, F. (1961). The Rate of Changes in the Rate of Articulation. In *Language and Speech*, 4:171-174.

Heldner, M. and Edlund, J. (2010). Pauses, gaps and overlaps in conversations. In *Phonetics*, 38:555-568.

Izre'el, S., Mello, H., Panunzi, A. & Raso, T. (2020). *In Search of Basic Units of Spoken Language. A corpus-driven approach*. Amsterdam: John Benjamin.

Lea, W.A. (1976). Prosodic Aids to Speech Recognition: 9. Acoustic Patterns in Selected English Phrase Structures

Martin, P. (2004). WinPitch Corpus: A text to Speech Alignment Tool for Multimodal Corpora. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04)*, Lisboa.

Moneglia, M. (2005). The C-Oral-Rom resource. In E. Cresti and M. Moneglia (Eds), *C-ORAL-ROM. Integrated reference corpora for spoken romance languages*. Amsterdam, John Benjamins:1-70.

Moneglia, M. and Raso, T. (2014). Notes on Language into Act Theory (L-AcT). In T. Raso and H. Mello (Eds), *Spoken corpora and linguistic studies*. Amsterdam: John Benjamins:468-495.

Moneglia, M. and Cresti. E. (2015). The cross-linguistic comparison of information patterning in spontaneous speech corpora: Data from C-ORAL-ROM ITALIAN and C-ORAL-BRASIL. In Klaeger, S. and Thörle (Eds.), *International Linguistics: Grammar and interaction in Romance Languages from a Contrasting Point of View*. Tübingen: Stauffenburg: 107-128.

Savy, R. (2005). Specifiche per la trascrizione ortografica annotata dei testi. In F. Albano Leoni and R. Giordano (Eds), *Italiano Parlato. Analisi di un dialogo*, Napoli, Liguori.