caletroscópio
Programa de Pós-Graduação em Letras: Estudos da Linguagem
ISSN (on-line) 2318-4574

Morphosyntactic, prosodic, functional and distributional description of the information unit of Parenthesis in spoken Brazilian Portuguese

Descrição morfossintática, prosódica, funcional e distribucional da Unidade Informacional de Parentético no português brasileiro falado

Saulo Mendes Santos[1]
Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil
Giulia Bossaglia[2]
Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

Resumo: Este tem por objetivo apresentar uma descrição preliminar da unidade informacional de Parentético (PAR), baseada no corpus C-ORAL-BRASIL (RASO; MELLO, 2012), representativo da fala informal da variedade mineira do português brasileiro, e embasada na Language into Act Theory (L-AcT; CRESTI, 2000). Foram pesquisadas as principais características prosódicas (duração, frequência fundamental, intensidade) e morfossintáticas (núcleos sintáticos, preenchedores mais frequentes), bem como características funcionais (metanarrativos, metalinguísticos e modais) e distribucionais (posição dentro do enunciado) desta específica unidade informacional. Nossos resultados sugerem que exista uma distinção funcional entre PAR curto vs. longo. Além disso, a unidade informacional de PAR parece, por vezes, carregar força ilocucionária.
Palavras-chave: Fala espontânea. Parentético. Português brasileiro. C-ORAL-BRASIL.

Abstract: This paper aims at presenting a preliminary description of the informational unit of Parenthesis (PAR), based on the corpus C-ORAL-BRASIL (RASO; MELLO, 2012), representative of the informal speech of the Minas Gerais's variety of Brazilian Portuguese, and underpinned on the Language into Act Theory (L-AcT; CRESTI, 2000). The main prosodic (length, fundamental frequency and intensity) and morphosyntactic (head types and most frequent fillers) characteristics, as well as functional (metanarrative, metalinguistic, and modal) and distributional (position within the utterance) characteristics were surveyed for this specific information unit. Our results suggest that a distinction between short and long PAR exist. Also, the information unit of PAR seems at times to carry illocutionary force.
Keywords: Spontaneous speech. Parenthesis. Brazilian Portuguese. C-ORAL-BRASIL.

1. Introduction

[1] Mestrando, Programa de Pós-graduação em Estudos Linguísticos, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil, saulo.mendes@gmail.com
[2] Professora Adjunta, Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil, giulia.bossaglia@gmail.com

As known from the handful of research undertaken over the past decades, spontaneous speech cannot be accounted for in the very same way that written language, due mainly to its peculiarities as a process and to phenomena entangled by the channel through which it is transmitted, say, the air. As an ongoing process carried out online by speakers in interaction, spontaneous speech is characterised by pauses, hesitations, repetitions, repairs, just to name a few. As a message put into code by the human phonatory system and transmitted throughout the surroundings of its source, spoken language has a fundamental frequency (f0), which enables the perception of the voice pitch, and an intensity, which causes the perception of volume, for each time unit pacing the speech stream. Since speakers dispose of different ways of conveying meaning, the structure of speech displays an organisation of the information different from written language. A good framework for a survey on the structure of speech is the one provided by the Language into Act Theory (L-Act), which is adopted for the purposes of this work. According to L-AcT, the Information Unit called Parenthesis corresponds, essentially, to a metalinguistic insertion that carries a speaker's evaluation about any part of the ongoing utterance. A chunk of the speech continuum is identified as a parenthetical insertion by means of the combination of three criteria: i) its prosodic characteristics; ii) its distribution inside the utterance; iii) its communicative function.

The information unit of Parenthesis (PAR) – or the corresponding units in other frameworks – has already been surveyed inside and outside the framework here adopted. The behaviour of this unit in spoken Brazilian Portuguese remains, however, to be better understood with respect to its functions, distribution, and prosodic properties. The purpose of this work is, thence, to take a few first steps into this subject by researching some of the main characteristics of PAR, as described on previous works, over a sample taken from a spoken Brazilian Portuguese corpus, representative mainly of the Minas Gerais State variety. The focus of this work lays on the description of PAR in the morphosyntactic, prosodic, functional and distributional levels. It must be acknowledged that the prosodic description reported on the following sections is of a very preliminary nature and that the most complex prosodic properties and patterns of PAR are set aside to a subsequent work. We tried, on the other hand, to grant some special attention to the morphosyntactic level by surveying the types of morphosyntactic content of PAR, as well as by crossing verb-headed PARs to some of their prosodic characteristics, say, mean f0 and f0 variation.

This paper is divided into four sections besides this introduction and the conclusion. The

second section reviews some basic assumptions and premises of the Language into Act Theory. The third section looks back at some previous works on functional and prosodic properties of PAR. The fourth section outlines the methodology of this survey and describes the sampled spoken corpus. The fifth section deals with the analysis of the data and its findings. Some final remarks are made in the conclusion, to sum the most important findings up and to point out what remains to be done.

2. Language into act theory (L-Act)

The L-Act (CRESTI, 2000; CRESTI & MONEGLIA, 2005; MONEGLIA & RASO, 2014) is inspired by the Theory of Speech Acts (AUSTIN, 1962), and it has been developed through extensive empirical research based on the LABLITA (*Laboratorio di Linguistica dell'Università di Firenze*, Italy) spoken corpora. L-AcT defines the theoretical and methodological foundations of the C-ORAL-ROM (CRESTI & MONEGLIA, 2005) and C-ORAL-BRASIL (RASO & MELLO, 2012) projects.

As an extension of the work left by Austin (1962), L-AcT assumes a pragmatic reference unit, say, the utterance, which is, on the contrary to other syntactic reference units, a more adequate unit for spontaneous speech. Furthermore, L-AcT introduces to the study of speech acts many prosodic notions which could only be captured in the light of technological tools developed during the past decades. Besides developing an innovative approach to speech acts, the research undertaken within this framework led to generalisations concerning the organisation of speech, which are of utmost importance for this work. Before discussing these generalisations, let us begin with some basic premises.

2.1 Utterance as a basic reference unit

The basic notions adopted in the framework of L-AcT emerge from the empirical research on spontaneous speech corpora, aiming at granting adequate categories to compile spoken corpora, which display particular characteristics compared to other types of corpora (MONEGLIA & RASO, 2014, p. 469). The transcription criteria within the C-ORAL family took into consideration the annotation of two types of prosodic breaks: terminal, annotated through double slashes "//", which delimit the utterance's boundaries, and non-terminal, annotated through a simple slash "/", delimiting tone units internal to the utterance

(MONEGLIA & RASO, 2014). According to Moneglia & Raso (2014, p. 469), the detection of prosodic breaks within the C-ORAL family corpora was assured by competent speakers, whose perception exhibits a high degree of convergence at cross-linguistic testing (MONEGLIA *et al.*, 2005; MONEGLIA *et al.*, 2010; RASO & MITTMANN, 2009; MELLO *et al.*, 2012).

Prosody plays a central role within L-Act. Besides proposing that prosodically terminated sequences delimit an utterance within the speech stream, L-AcT assumes that this speech chunk encloses the prosodic cues which, together with other pragmatic-cognitive parameters, convey a specific illocutionary force to the utterance. L-AcT defines the utterance as a prosodically terminated sequence within the speech flow endowed with a specific illocutionary force conveyed by specific prosodic features together with other contextual parameters. In other words, an utterance is a pragmatically interpretable speech chunk (CRESTI, 1994; CRESTI, 2000; MONEGLIA & CRESTI, 2006; MONEGLIA, 2005, 2006, 2011) bounded by terminal prosodic breaks.

2.2 Information Patterning Theory

Within L-AcT framework, an utterance may be simple, i.e. formed by a single tone unit, or complex, i.e. internally segmented by functional (not created by hesitations and repairs) non-terminal prosodic breaks, thus formed by more than one tone unit and, then, displaying a prosodic structure, i.e., a prosodic pattern. According to Moneglia & Raso (2014), the prosodic pattern set up by these breaks is, in broad lines, isomorphic to the information patterning, that has traditionally been described by different frameworks in terms of Topic-Comment, Topic-Focus, Theme-Rheme, or Given-New (CHAFE, 1970; CHOMSKY, 1971; HALLIDAY, 1976; LAMBRECHT, 1994; HOCKETT, 1958; JACKENDOFF, 1972; KRIFKA & MUSAN, 2012). Consistent with the needs emerging from the research undertaken on the C-ORAL corpora, L-AcT went beyond this opposition and adopted a rich set of new information units known as the Information Pattern Theory (CRESTI, 1994; MONEGLIA; CRESTI, 2006; CRESTI; MONEGLIA, 2010; MELLO; PANUNZI; RASO, 2011).

According to Moneglia & Raso (2014, p. 479), the prosodic model of the Information Patterning Theory is underpinned on the IPO model (‘T HART et al., 1990). The IPO model puts forth that functional prosodic cues, such as relevant, voluntary f0 movements, can convey a specific informational value. The model was later expanded at LABLITA (FIRENZUOLI,

2003) in order to encompass other prosodic profiles defining other information units.

According to the Information Patterning Theory, the information units can be of two main types, say, textual units and dialogic units. The former makes up together with COM, the sole information unit that is indeed necessary to build an utterancem the semantic and syntactic content of the utterance, while the latter serves basically to regulate the ongoing interaction. Below, we summarise the information units set up by the Information Patterning Model:

Table 1 – Information units according to the Information Patterning Theory

	Information unit	Tag	Function
Textual units	Comment	COM	Conveys the illocutionary force of the utterance
	Topic	TOP	Determines the scope of application of the illocutionary force conveyed by COM
	Appendix of Comment	APC	Combines with the text of COM and concludes and utterance
	Appendix of Topic	APT	Delivers delayed information to the text of TOP
	Parenthesis	PAR	Inserts metalinguistic and metanarrative information into the utterance with backward or forward scope, help the addressee to understand what is being said
	Locutive Introducer	INT	Signals that the subsequent units have a different status regarding what was being said, marking a shift in the interpretation
	Multiple Commnet	CMM	Build up a chain of COM which sets an illocutionary pattern
	Bound Comment	COB	Build up a sequence of COM which does not set an illocutionary pattern rather being the expression of the flow of thought
Dialogic units	Incipit	INP	Opens the communicative channel
	Conative	CNT	Requests an interactant to join or to refrain any deed
	Phatic	PHA	Keeps the communicative channel open
	Allocutive	ALL	Pinpoints the interactant to whom the utterance is addressed
	Expressive	EXP	Provides emotional support for the interaction indicating affinity and empathy
	Discourse Connector	DCT	Governs the discourse flow connecting information units and showing that a sequence is not yet finished

(adapted from MONEGLIA & RASO, 2014, p. 490-491)

An information unit is thus said to have three defining components: i) a prosodic profile; ii) a distribution, namely its position with respect to the illocutionary unit; and iii) a specific pragmatic function, such as the ones displayed in Table 1.

Insofar as utterances are internally structured through the combination of information units, two types of syntactic configuration are envisaged within L-AcT: *linearised* and *patterned* (CRESTI, 2014). On the one hand, the syntactic relations observed within a single information unit (linearised syntax) are understood as proper subordination and coordination

structures. On the other hand, syntactic constructions performed across more than one information unit are part of a patterned construction and, thus, in principle, not subject to proper syntactic relations. Therefore, L-AcT assumes that, although syntactic relations may be observed between information units, what prevails are the informational/pragmatic relations between them, conveyed by prosodic cues. All the same, this statement is disputed.

3. Information unit patenthesis (PAR)

The information unit of Parenthesis is functionally described, within the L-AcT, as an insertion of semantic information – that is to say, compositional regarding the utterance – inside the utterance. Basically, PAR somehow restricts or clarifies the use of any term or expression as well as other facts narrated by the speaker, having either backward or forward scope (TUCCI, 2010; MONEGLIA & RASO, 2014).

There have also been relevant contributions to this subject outside the L-AcT framework. Noteworthy is the handful of research carried out by Schneider (2014), who has been studying in depth the functions that parentheses have cross-linguistically, based on spoken corpora. This author defines parenthesis as a particular linguistic behaviour or communicative strategy brought about by the speaker's need for inserting additional information outside the level of the ongoing utterance, thus breaking its linearity. In this sense, the parenthesis is marked by a syntactic, semantic or prosodic disruption, or else a combination of disruptions on one or more of such levels (SCHNEIDER, 2014). According to Schneider (2014, p. 287), many other functions have been hypothesised for parenthesis. To sum them up, parenthesis may fulfil functions such as: side notes aiming to correct interferences or hindering reactions of the interactant (BERRENDONNER, 2010); specifications, exemplifications, explications, descriptions or definitions of referential units (MAZELAND, 2007); information structure units (TAGLICHT 1984; BRANDT, 1996; ZIV, 2002; KALTENBÖCK 2010); epistemic extenuation of speaker commitment (SCHNEIDER, 2007); epistemic increase of speaker commitment (KALTENBÖCK, 2010); reported speech (SCHNEIDER, 2007); evidential information (IFANTIDOU, 2001; SCHNEIDER, 2007); indication of illocutionary force (SCHNEIDER, 2007: SCHNEIDER, 2010); self-initiated repair (SCHNEIDER, 2007; BERRENDONNER, 2010; SCHNEIDER, 2011); resumption of something precedingly said (SCHNEIDER 2011); focusing (SCHNEIDER, 2007); hesitation (SCHNEIDER, 2007); and

conative or phatic function (SCHNEIDER, 2007). The overlapping between many of the functions set forth for parenthesis and the functions of some information units within L-AcT framework – such as conative, phatic, focusing – is striking. L-AcT deals with many of them not as parentheses, but as particular information units characterised by specific prosodic profiles and labelled with different names (Conative, Phatic, among others: MONEGLIA & RASO, 2014).

An information unit is defined in terms of functional, prosodic and distributional criteria. As for the data of the Italian section of C-ORAL-ROM, Tucci (2010) reports the main functions and possible distribution for PAR as well as general prosodic characteristics. Firstly, PAR is described in terms of three main functions, namely a metanarrative, a modal, and a metalinguistic one. Metanarrative PAR inserts information to instruct the addressee about the speaker's either positive or negative attitude towards the narrative. Modal PAR, on the other hand, inserts the speaker's judgment – a hedging or any other modality (epistemic, deontic and so forth) – about the utterance content. Lastly, metalinguistic PAR signals to the addressee that a particular and subjective lexical selection was done, frequently explaining such selection by reformulating it.

Tucci (2010) also describes the prosodic profile for PAR based on data from Italian. According to the author (TUCCI, 2010, p. 8), PAR is prosodically characterised in comparison to the neighbouring units by i) a sensible descent of mean f0 values on the onset; ii) an overall flat, deep f0 movement along the prosodic unit; iii) change in speech rate; iv) possibility of emphatic peaks. Its prosodic profile has thence the following features: i) absence of preparation; ii) flat or descending nucleus; iii) possibility of ascending coda within the utterance; iv) mean duration of nuclear syllable: 0.142 ms; nuclear f0 excursion[3] between 104 Hz-90 Hz (male) or 134 Hz-115 Hz (female); onset values: 104 Hz (male), 134 Hz (female), maximum: 173 Hz (male), 157 Hz (female), minimum: 96 Hz (male), 127 Hz (female). Finally, Tucci (2010) reports that PAR has a free distribution within the utterance, except for the initial position.

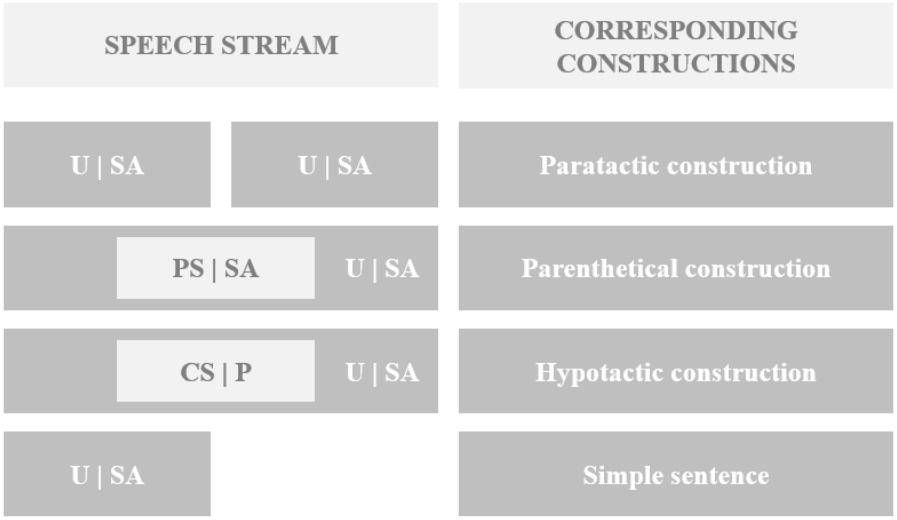
Tucci (2010) considers PAR a borderline structure with respect to syntax. On the one hand, PAR is part of the linear syntax of the host utterance, since it combines with other syntactic structures within the utterance. On the other hand, PAR is not commanded by the syntactic relations of its host utterance, since its order is not given by the same principles that

[3] Excursion rate corresponds to f0 final – f0 initial.

control, for instance, government. Therefore, PAR has no syntactic relation, neither hypotactic nor paratactic, with the host utterance, being thus a linearly adjacent structure but not a constituent (TUCCI, 2010). Likewise, Tucci (2010, p. 3) points out that the independence of PAR regarding its host structure correlates with its prosodic independence.

Schneider (2014) distinguishes *parenthesis* from *parenthetical sequence*. According to the author, the latter is a particular kind of linguistic behaviour or communicative strategy, whereas the former is a concrete item producing a constructional pattern. Schneider (2014) puts forth that at least one type of parenthetical sequence is a speech act combined with another utterance, also bearing a speech act. The author also distinguishes three constructions: a paratactic construction, a parenthetical construction, and a hypotactic construction. We adapted below the scheme proposed by Schneider (2014, p. 283):

Figure 3 – Parenthetical constructions; adapted from Schneider (2014, p. 283)



In Figure 3, each construction may be an Utterance (U), a Parenthetical Sequence (PS) or a Clausal Sequence (CS). Each construction bears either a Speech Act (SA), or a Proposition (P) without an illocutionary force. Paratactic constructions are merely adjacent utterances, both carrying illocutionary force. A parenthetical construction, on the other hand, is one in which both host utterance and the insertion have illocutionary force. In the hypotactic construction, the insertion does not have any illocutionary force. The parenthetical sequence bearing illocutionary force is provisionally called by the author a *side note*.

4. Data and methodology

In this section we outline our methodology to describe Parenthesis.

4.1 Data

The survey on PAR was based on a sample of the C-ORAL-BRASIL corpus (RASO & MELLO, 2012), a spontaneous speech corpus representative of the Minas Gerais State diatopy of Brazilian Portuguese, characterised by a wide diaphasic variation.

The corpora of the C-ORAL family have characteristics specially designed to enable an adequate study of spontaneous speech, such as: a) the size of the corpora, whose total number of words varies between 300,000 and 450,000; b) high degree of diaphasic variation, which is paramount in order to document different types of illocutions; c) attention to the diastratic variation; d) coverage of varied types of interaction, namely monologues, dialogues (two active participants), and conversations (more than two active participants); e) orthographically based transcription, enabling computerised processing without limiting the documentation of ongoing processes of grammaticalisation and lexicalisation; f) prosodic breaks' segmentation, and annotation of interruptions and repairs; g) text-to-speech alignment (through the software WinPitch, MARTIN, 2005); h) morphosyntactic parsing; i) high-quality audio files, recorded through non-invasive devices, in order to preserve the spontaneity of the natural-context interactions (see MELLO 2014 for more details).

The sample we used is the Brazilian Portuguese minicorpus extracted from the C-ORAL-BRASIL, available online for corpus queries at the DB-IPIC (DataBase for Information Patterning Interlinguistic Comparison) website (http://www.lablita.it/app/dbipic/) developed by the LABLITA. According to Panunzi & Mittman (2014, p. 130), the minicorpus contains a total of 20 files taken from C-ORAL-BRASIL. At the same time as it is made up by the best audio quality recordings, the minicorpus preserves the essential architecture of the C-ORAL-BRASIL corpus. The minicorpus is provided with a manual informational tagging, that allows a quick identification of the information units, according to the Information Patterning Model of L-AcT.

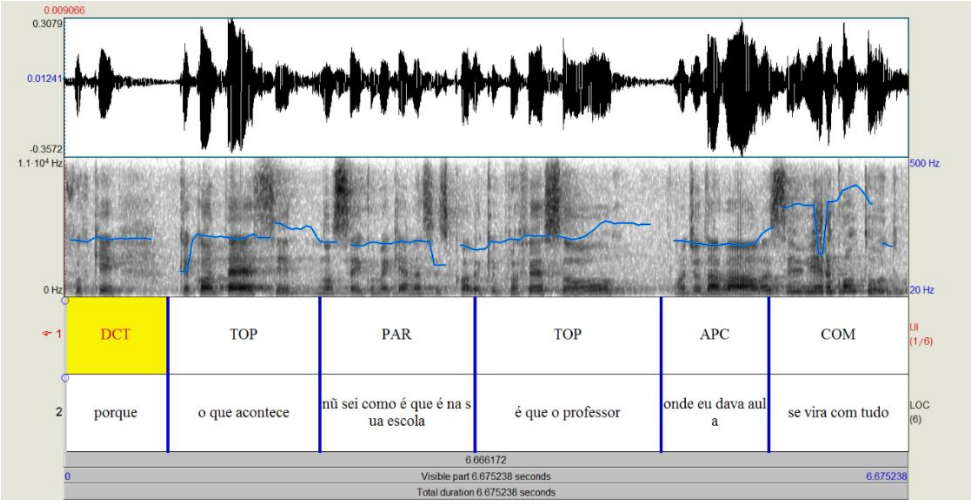
4.2 Methodology

All occurrences of PAR found in the minicorpus were sampled for this work. We surveyed the sample for: i) morphosyntactic properties (dependency relations, syntactic head

type); ii) functional typology; iii) duration; iv) distribution; and v) presence of illocutionary force.

Furthermore, we annotated all information units of 10 utterances randomly chosen from the sample which contained at least one PAR, and extracted basic prosodic measures (mean f0, intensity), using Praat (BOERSMA & WEENINK, 2017), in order to sketch a prosodic description. Additionally, we annotated all verb-headed PARs and extracted their basic prosodic measures (f0, intensity, f0 excursion). An example of the annotation made is displayed below:

Figure 4 – Example of annotation



The data was thence classified through the audit of the audios, using the software Winpitch (MARTIN, 2005). In the next section, we set out the tabulated data and our analyses.

5. Results and analysis

Insofar as there are more and less prototypical occurrences, we illustrate each category with some of their most representative examples in order to bring forward the guidelines that led the categorisation process.

In our sample, we found a total of 152 tokens of PAR. A total of 68 PAR tokens displayed a verbal head and were, therefore, annotated for prosodic measurement. Most of PAR occurrences – 107 tokens – were retrieved from monologues, which typically display more complex utterances.

5.1 PAR length

In order to describe PAR length, we classified as *short* all tokens corresponding to a single phonological word, say, containing a unique primary stress, and as *long* all other tokens:

(2) a) short

<oito /=CMM= né /=CMM= na verdade> //=PAR= (audio bfamdl01_296)

<eight /=CMM= huh /=CMM= *actually*> *//=PAR=*

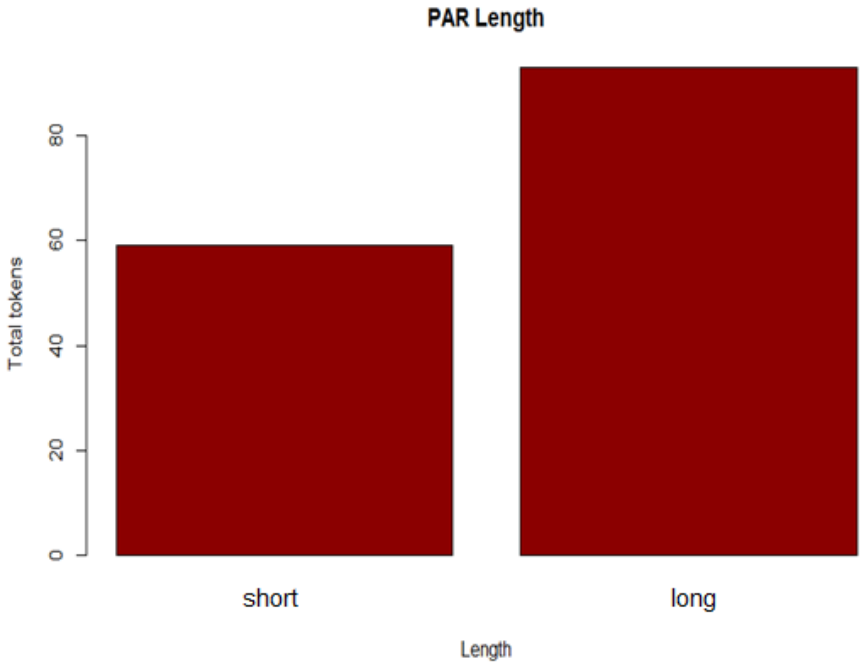
b) long

fulana de tal /=TOP_r= à /=SCA= Casa Dragão /=SCA_r= deve /=COM_r= Casa Dragão é minha casa //=PAR= (áudio bfammn03_123)

Janie Doe /=TOP_r= to /=SCA= Casa Dragão /=SCA_r= owes /=COM_r= *Casa Dragão is my store //=PAR=*

Out of 152, 93 tokens were classified as long. Chart 1 below displays the distribution of PAR with respect to length:

Chart 1 – PAR length distribution



The distinction between short and long is of utmost importance for two main reasons. Firstly, fillers of short PAR are most of the times limited to very few possibilities, as shown in subsection 5.5. Secondly, their functions also seem to be more restricted, as shown in subsection 5.6. This leads to the conclusion that some PAR, such as *na verdade* (actually) in (2), may be in process of grammaticalisation or discursivisation. In accordance to what has been put forth by other authors (see SCHNEIDER, 2007; MAZELAND, 2007), we consider that our data provide additional evidence for granting short PAR tokens a peculiar status.

5.2 Position within the utterance

To classify our data, we took into account three possibilities, namely *medial*, *final*, and *embedded* within another informational unit. Examples in (3) illustrate these possibilities:

(3) a) medial

<eu queria ver a comunidade dele> lá /=COB= ver que que es falam /=SCA= entre si /=COB= assim /=PAR= só pra eles mesmos //=COM= (audio bfamcv01_27)

<I would like to know his community> there /=COB= to hear what they talk /=SCA= to each other /=COB= *like /=PAR=* just for themselves //=COM=

b) final

<com> hífen /=COM= por exemplo //=PAR= (audio bfamcv04_284)

<with> hyphen /=COM= *for instance //=PAR=*

c) embedded

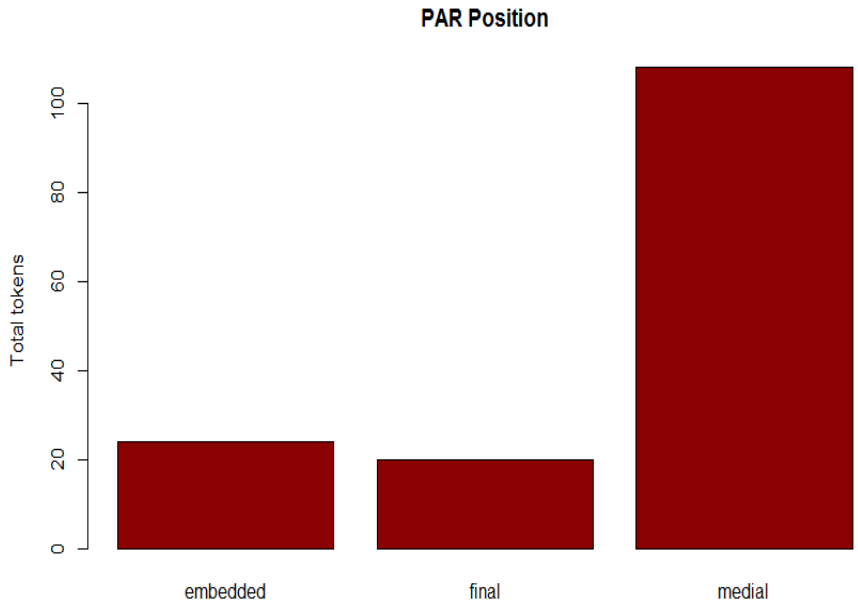
aí cê /=i-COM=[4] *tipo /=PAR= dá a &irre [/2]=SCA= a real de quantas palavras que são //=COM=* (audio bfamcv04_277)

then you /=i-COM= *sort of /=PAR=* say the &tru [/2]=SCA= how many words there are truly //=COM=

[4] Interrupted information units are tagged with "i-". PAR can occur within another information unit, as in this example.

The position distribution of PAR is set out in Chart 2 below:

Chart 2 – PAR position distribution



It is noteworthy that most PAR tokens are embedded and medial. The fact that final position PAR may distributional and functionally overlap information units of Appendix of Comment helps to explain why this position is disfavoured. Conversely, the embedded position seems to be very characteristic of PAR to the extent that textual information units are not expected to be intertwined with other information units.

5.3 Syntactic head types

The syntactic classification took into consideration the main phrase types plus a subclassification of verb phrases into *sentential*, *finite*, and *non-finite*. For the sake of economy, we exemplify only the verbal phrases:

(4) a) Sentential
nesse mercado eu fiquei um bom tempo /=COB= *eu já tinha tido uma experiência* /=*PAR*= e é um mercado muito carente de informação //=COM= (audio bfammn06_54)

I worked for a long time in this sector /=COB= *I had already had an experience /=PAR=* and this is a sector where information lacks a lot //=COM=

b) Finite
que a Dodora /=i-COB= que era a mamãe /=PAR= ia ficar < muito > feliz /=SCA= de morar /=SCA= perto da dona Terezinha /=COB= que era minha &b [/2]=SCA= minha avó //=PAR= (audio bfammn02_143)

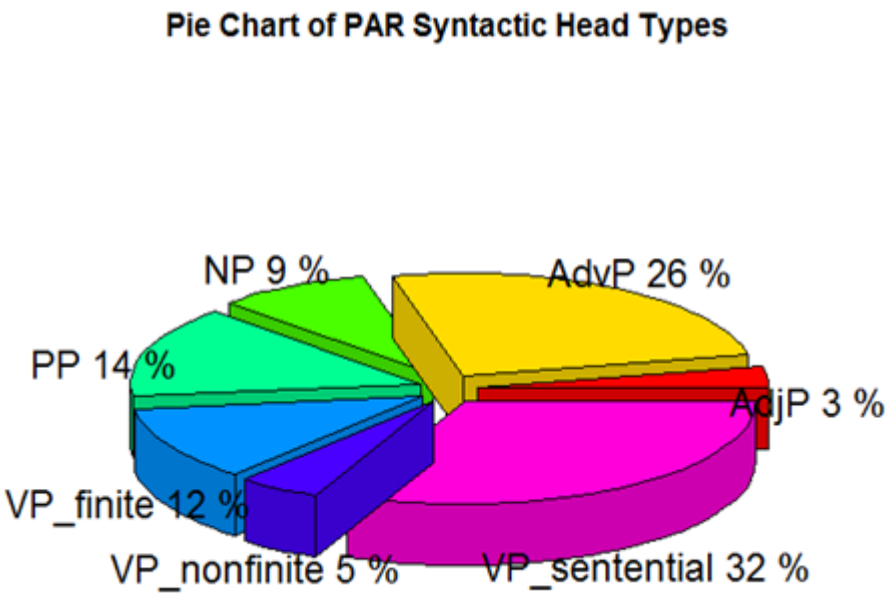
that Dodora /=i-COB= who was my mother /=PAR= would be < very > happy /=SCA= to live /=SCA= near Dona Terezinha /=COB= *who was my &b [/2]=SCA= my grandmother //=PAR=*

c) Non-finite
se a gente vai de [/1]=SCA= de táxi /=TOP= voltar de táxi /=PAR= po' comprar um //=COM= (audio bfamdl01_201)

if we go [/1]=SCA= by cab /=TOP= *come back by cab /=PAR=* you could buy one //=COM=

Chart 3 below shows PAR head type distribution:

Chart 3 – PAR head type distribution



As it can be seen from the chart, verbal PAR – 49% – is by far the most frequent type. Among other types, adverbial PAR is the most frequent – 26%. It is noteworthy that occurrences of PP, NP, AdvP, and AdjP are mostly short PAR tokens. On the other hand, verbal PAR tokens are mostly long, as expected.

5.4 Formal syntactic relation

Even though L-AcT framework considers that there are no linear (or proper) syntactic relations across information units, we surveyed for syntactic dependency relations established between PARs and neighbouring units. We classified tokens as *dependent* or *independent*, as illustrated in (5):

(5) a) Dependent

< dá licença um > pouquinho /=COB= enquanto cê nũ tá jogando /=PAR= que cê [/1]=SCA= ninguém güenta esse cu seu não //=COM= (audio bfamcv03_243)

< excuse me for a > moment /=COB= *while you're not playing /=PAR=* that you [/1]=SCA= nobody puts up with this ass of yours //=COM=

b) Independent

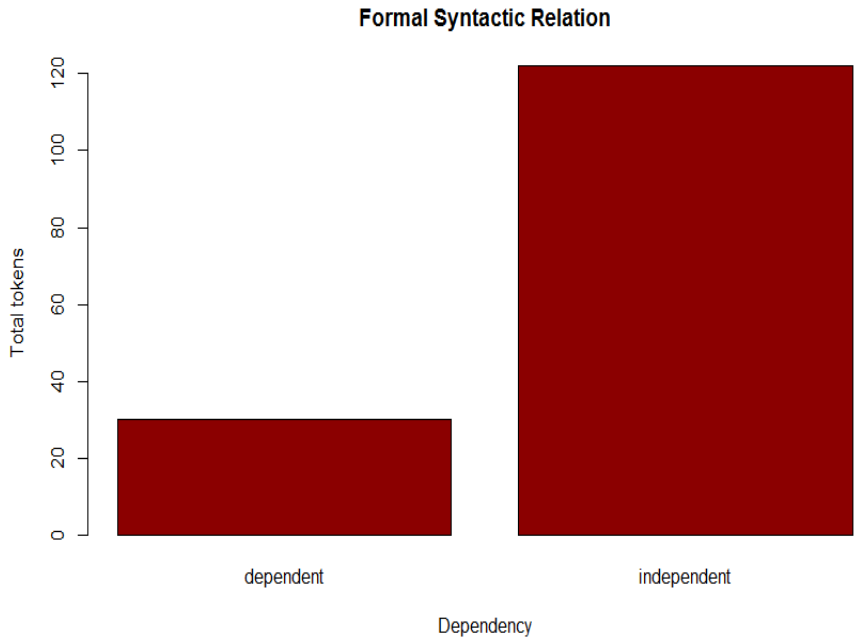
que /=DCT= quando nós fomos levar o papel da advogada lá pra assinar /=TOP= que

a advogada é que mexeu pra mim /=PAR= ela nũ queria assinar //=COM= (audio bfammn05_87)

that /=DCT= when we took the attorney's document for her to sign /=TOP= *that there is an attorney championing my cause /=PAR=* she didn't want to sign it //=COM=

Although "b" exhibits what appears to be a complementiser, it cannot be deemed dependent on the structure of the host utterance, since it does not establish a hypotactic relation to it. It is neither a relative nor a subordinate clause. Rather, it inserts an explanation in order to avoid a reaction by the addressee, who was not aware that there was an attorney championing the cause. The result is summarised in Chart 4 below:

Chart 4 – PAR head type distribution



The chart above displays that most PAR tokens exhibit no true syntactic relation of dependency to its host utterance anyway. This figure is of paramount importance because it brings forward the fact that PAR undergoes not only a prosodic disruption but also frequently a syntactic disruption with respect to its host utterance.

5.5 Lexical fillers

Aiming to provide additional evidence for a possible ongoing grammaticalisation or discursivisation process among some PAR types, we surveyed the most frequent fillers and collocations. Therefore, we searched for the most frequent trigrams, bigrams, and single words recurring within the text of PAR. Table 2 below displays the most frequent words and expressions (functional lexemes included) collapsed for short and long PAR:

Table 2 – Most frequent fillers within PAR

short PAR		long PAR	
Filler	**Occurrences**	**Filler**	**Occurrences**
assim	23	que	33
por exemplo	8	de	15
tipo	8	é	12
eu acho	4	da	11
por aí	2	eu	11
aqui	1	o	11

As observed in the corpus, long PAR tokens carry various types of fillers and exhibit no preference for any pattern with respect to words or collocations, therefore, showing no evidence for processes of grammaticalisation. As expected, among more frequent occurrences they display only functional words. Conversely, 45 out of 59 short PAR tokens display recurrent full lexical items or collocations, such as *por exemplo* (for instance) and *eu acho* (I think), exhibiting a strong pattern towards modal and discursive functions. We regard this fact as a *prima facie* evidence that short PAR must not have the same status as long PAR, and that some of them may be undertaking a grammaticalisation or a discursivisation process, undergoing desemanticisation, decategorialisation, and even phonetic loss. We also consider that some tokens classified in the corpus as short PAR might functionally overlap other information units (if they could not be better classified under such labels).

5.6 Functions of PAR

We classified PAR into the three main functions reported by Tucci (2010): metanarrative, metalinguistic and modal. During the analysis, we also observed some tokens

whose function was to pinpoint any object in the context or a gesture probably occurring simultaneously to the utterance. Although there were very few tokens of this sort, we deemed them noteworthy, since we suppose they may establish direct relations between information units and their contexts[5]. Examples of the functions are shown in (6):

(6) a) metanarrative

e /=DCT= e o seu Pedro /=TOP= o marido dela também /=PAR= muito bravo também /=PAR= &s [/1]=EMP= sistemático /=PAR= um dia ele /=TOP= já depois do quase setenta ano /=TOP= resolveu /=SCA= &a [/1]=EMP= arranjar outra //=COM= (audio bfammn03_4)

and /=DCT= and Mr. Pedro /=TOP= *also her husband /=PAR= also a very sulky man /=PAR= &s [/1]=EMP= methodical /=PAR=* one day he /=TOP= already over 70 years old /=TOP= decided /=SCA= &a [/1]=EMP= to find another woman //=COM=

b) metalinguistic

bom /=INP= aí a bolsa chega /=CMB= aí /=DCT= a gente /=SCA= homogeniza /=COB= desgruda as plaquetinhas da parede < dela assim > /=PAR= faz tipo uma /=SCA= cinturinha +=COB= (audio bpubcv01_116)

well /=INP= so the blood bag arrives /=CMB= and then /=DCT= we /=SCA= homogenise it /=COB= *we take off the platelets from the inner sides of the bag < like that > /=PAR=* we do sort of a /=SCA= waist +=COB=

c) modal

só que é de microondas /=COM= eu acho //=PAR= (audio bfamdl01_40)
but that's for microwaves /=COM= *I think //=PAR=*

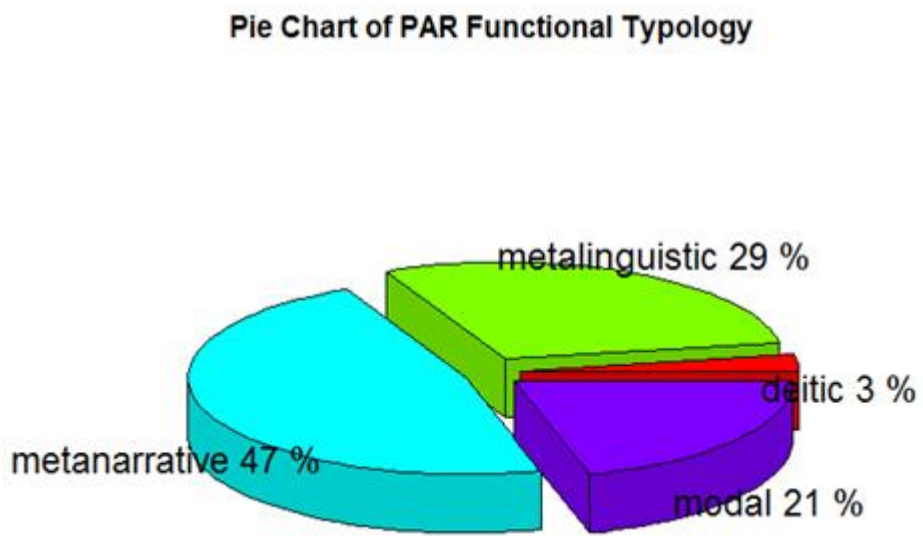
d) deictic

[5] As a matter of fact, C-ORAL-BRASIL corpus consists of text-to-audio aligned files, thus only allowing a conjecture of such a connection. A survey in a multimodal corpus would be necessary to validate this conjecture.

< *e uma* > *dessa /=COM= aqui /=PAR= também* < *o'* > *//=APC=* (audio bpubdl02_41)
< and one > of these /=COM= *here /=PAR=* < look > //=APC=

Chart 5 below presents the overall distribution of PAR into the four subfunctions:

Chart 5 – PAR functional typology



As seen in the chart, most PAR tokens, 47%, are metanarrative, followed by metalinguistic and modal PAR. However, this overall figure does not seem to say anything if we do not collapse it by length. Charts 6 and 7 do this work:

Chart 6 – Functions of short PAR

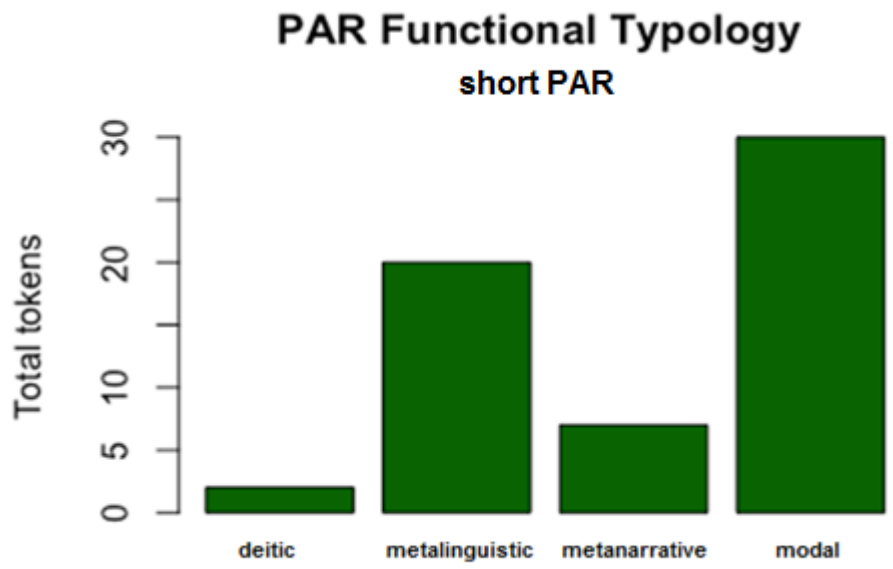
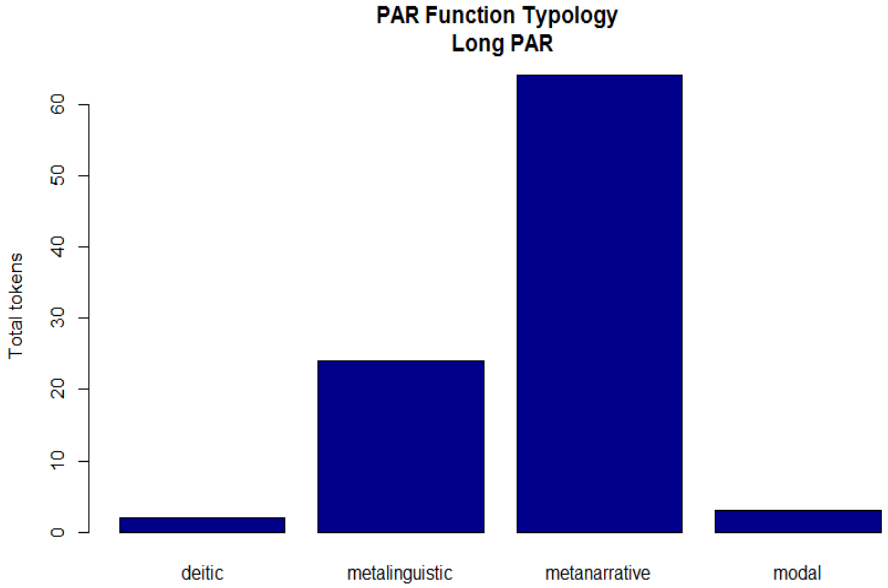


Chart 7 – Functions of long PAR

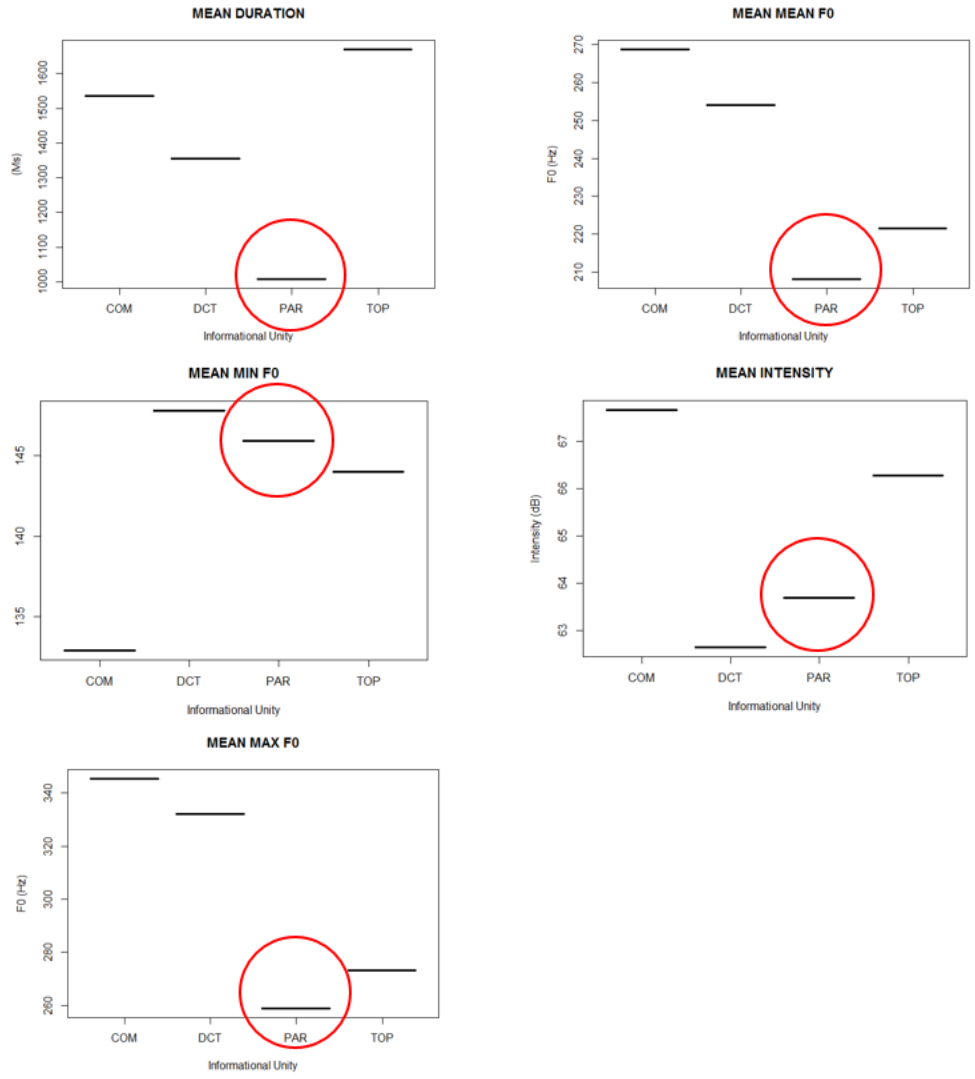


We can now observe a clearer picture in comparing charts 6 and 7. Whereas short PAR tokens display a slope towards modal and metalinguistic functions, long PAR tokens tend to be mostly metanarrative and metalinguistic. However, this distinction could become even clearer provided that many short PAR tokens classified as metalinguistic can be considered a kind of grammaticalised/discursivised PAR, or even reclassified into another information unit.

5.7 Initial prosodic description

In order to have an overview of prosodic characteristics of PAR regarding its neighbouring units, we annotated all information units found within 10 randomly taken utterances containing PAR in our sample. We measured mean duration, mean mean f0, mean maximum f0, mean minimum f0, and mean intensity, using the Praat script Prosody Pro (XI, 2013). The means for each label were calculated and are shown below in Chart 8:

Chart 8 – Prosodic means of PAR and neighbouring units DCT, TOP, COM



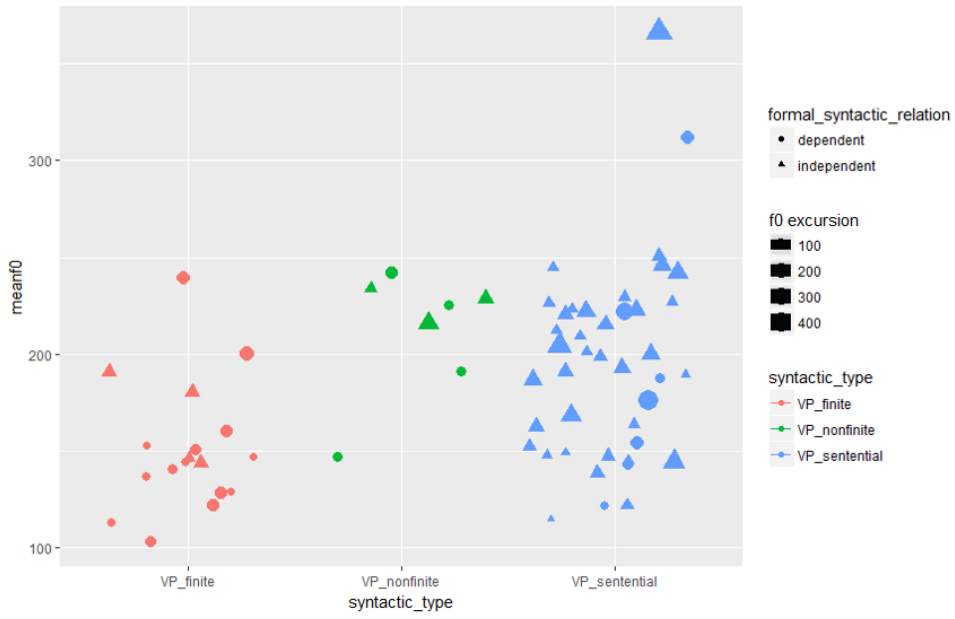
In Chart 8, COM stands for Comment, DCT for Discourse Connector, PAR for Parenthesis, and TOP for Topic (see Table 1). PAR means are circled in red. As regards mean duration, PAR is the shortest unit among the units surveyed, probably due to an influence of short PAR tokens,

since the difference between short and long PAR was not considered here. As reported for Italian, the mean mean f0 of PAR is also the lowest among the textual units surveyed, thus providing *prima facie* evidence that in Brazilian Portuguese f0 movement of PAR also displays a negative change compared to the neighbouring units (remember that PAR was described as having a flat, deep f0 contour in Italian). In accordance with this picture, the mean minimum f0 of PAR is among the highest values and mean maximum f0 is among the lowest, showing that PAR probably does not display in general a very ample f0 range. For instance, whereas COM displays a mean range of approx. 212Hz, PAR exhibits a mean range of approx. 115Hz, that is, a difference of more than 100Hz, which is indeed prosodically relevant. Finally, the mean intensity of PAR also displays a tiny difference regarding other textual units, though we cannot deem it relevant for reasons going beyond the scope of this work.

5.7 f0 measures for verb-headed PAR

In this subsection, we present some f0 measures taken from verb-headed PAR, collapsed into finite, non-finite, and sentential PAR. Chart 9 below sets out the results:

Chart 9 – f0 measures vs. verb-headed PAR



If little can be said about non-finite verbal PAR (green), some differences between finite verbal (red) and sentential PAR (blue) can be drawn out. Firstly, finite PAR tokens are scattered in lower f0 area of the plot, thus displaying a tendency toward a lower mean f0. On the other hand,

sentential PAR tokens are scattered mainly between 200 and 250 Hz, that is comparable to COM (250 Hz, Chart 5). Secondly, most finite PAR tokens are syntactically dependent (circle), which was indeed expected, whereas most sentential PAR are independent (triangle). Lastly, sentential PAR tokens display a broader excursion – max f0 - min f0 – (dot size), whereas finite PAR tokens have, aside from a few tokens, a narrower f0 range. If any PAR is to exhibit illocutionary force, we should expect this to happen mainly amid sentential PARs, since they present not only wider f0 excursion and higher f0 mean, but also a syntactic – probably semantic and discursive – disruption to its host utterance.

5.8 Illocutionary PAR

Even though it is a tenet within the L-AcT framework that only three information units, say, COM, CMM and COB, are entitled to bear illocutionary force, we have surveyed our sample in order to find possibly illocutionary PAR tokens insofar as Schneider (2014) reports parenthetical sequences bearing illocutionary force. However, the lack of criteria to classify illocutionary and non-illocutionary information units hampered an overall picture. All the same, there seems indeed to exist PAR with some illocutionary force. Below follows an example:

(7) a) Illocutionary PAR

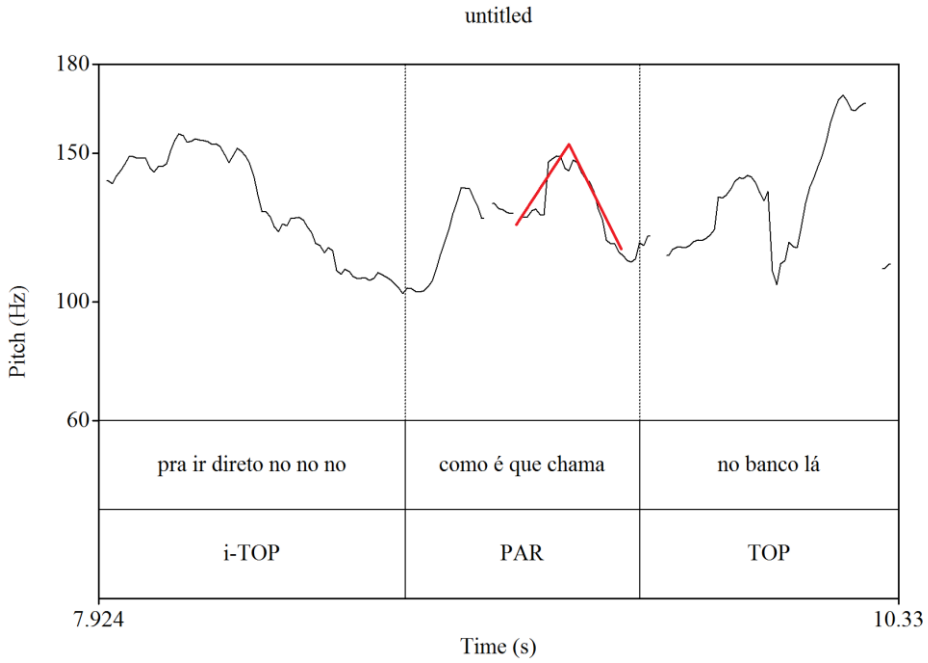
hhh agora /=PHA= &t [/1]=EMP= essa conta da caixa /=TOP= Geraldo falou comigo que o [/1]=SCA= o [/1]=EMP= o camarada tava aqui antes /=TOP= < tava > numa [/2]=EMP= não /=PHA= &he /=TMT= pra ir direto no [/1]=i-TOP= no [/1]=EMP= no [/1]=EMP= como é que chama /=PAR= no [/1]=EMP= no banco lá /=TOP= é direto lá //=COM= (audio bpubcv02_97)

hhh now /=PHA= &e [/1]=EMP= this bill of caixa /=TOP= Geraldo told me that [/1]=SCA= a [/1]=EMP= a man was here /=TOP= < he was > in a [/2]=EMP= no /=PHA= &he /=TMT= to go straightforward to [/1]=i-TOP= to [/1]=EMP= to [/1]=EMP= *what's the name /=PAR=* to [/1]=EMP= to that bank /=TOP= straightforward there //=COM=

This token displays many characteristics that lead us to think of an illocutionary PAR token. In this case, PAR is embedded within a TOP token, which, as we have shown, is a typical position

for PAR to occur. Therefore, the fact that this token is indeed a PAR seems indisputable. Furthermore, the token shows a long metanarrative PAR preceded by hesitations and repetitions, aiming to hint the addressee that the speaker does not remember the name of the bank. Figure 5 shows its f0 contour:

Figure 5 – f0 contour of example (7)



On the left side, we can observe part of the preparation of TOP plus some not transcribed hesitations. On the right side, we can observe a final ascending f0 contour, typical to one of the three types of TOP (CAVALCANTE, 2015). PAR, in the centre, is embedded within a TOP unit. Notwithstanding the not very prominent f0 peak, which may be explained by the speaker's need for conveying a PAR function by lowering and flattening f0 contour, the prosodic profile of PAR seems to display the characteristic profile of a total question, marked by the red lines. This kind of PAR resembles, to some extent, the type of parenthetical sequence carrying a speech act, say, a side note, as proposed by Schneider (2014).

6. Conclusion

In this paper, we have surveyed some initial properties of the information unit of Parenthesis, as it is defined within L-AcT, based on a representative sample of the C-ORAL-BRASIL corpus. As a result, we found that the basic prosodic measurements carried out on a

part of our sample seem to indicate that Parenthesis in Brazilian Portuguese behaves in broad lines in the same way as in Italian, say, with a deeper f0 contour with respect to its neighbouring units. Additionally, when a Parenthesis is inserted within an utterance, not only a prosodic break but also a syntactic disruption is frequently observed. As aforesaid, this is by no means a definitive work, since many issues remain to be deepened. All the same, this survey led us to some preliminary conclusions. Firstly, short parentheses deserve to be granted a special status, since their fillers display an ongoing grammaticalisation/discursivisation process, and since they display a slope towards the modal function. Furthermore, some tokens, especially short metalinguistic PAR, need to be better studied, to ascertain whether they can better fit within the group of *dialogic* information units. We also observed that long sentential PAR tokens show a wider f0 excursion, indicating that, if there is to be an illocutionary PAR, it might be amid this type. Finally, though we still have not been able to deal with the existence of illocutionary PAR, we provided an example of a token apparently bearing an illocution. Yet, much remains to be done. A more accurate, in-depth and extensive prosodic description of PAR could shed some light not only on the issue of illocutionary force, but also on the parenthesis profile and its plethora of functions. To do so, many fine-grained details should be dealt with, such as gender of speakers, position within the utterance, functions, possible prosodic nuclei, speech rate, f0 contours and movements, just to name a few.

Referências

AUSTIN, John L. *How to do things with words*. London, Oxford University Press, 1962.

BERRENDONNER, Alain. Pour une praxéologie des parenthèses. In: CORMINBOEUF, Gilles; HEYNA, Franziska; AVANZI, Mathieu (Eds.). *Les parenthèses en français*. Nancy, Presses Universitaires de Nancy, 2010, p. 5–23.

BOERSMA, P.; WEENINK, D. *Praat: doing phonetics by computer* [Computer program]. Version 6.0.33, 2017, Disponível em http://www.praat.org/.

BRANDT, Margareta. Subordination und Parenthese als Mittel der Informationsstrukturierung in Texten. In: MOTSCH, Wolfgang (ed.). *Ebenen der Textstruktur: Sprachliche und kommunikative Prinzipien*. Tübingen, Niemeyer, 1996, p. 211–240.

CAVALCANTE, Frederico. *The topic unit in spontaneous american english: a corpus-based study*, 2015, 184 fls., (Dissertação de mestrado em Linguística Teórica e Descritiva) Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte.

CHAFE, Wallace L. *Meaning and the Structure of Language*, Chicago, The University of Chicago Press, 1970.

CHOMSKY, Noam. Deep structure, surface structure and semantic interpretation. In: *Semantics*: An Interdisciplinary Reader. STEINBERG, Danny D.; JACOBOVITS, Leon A. (eds). Cambridge, Cambridge University Press, 1971, p. 183-216.

CRESTI, Emanuela. Speech act units and informational units. Speech Acts and Linguistic Research, Proceedings of the Workshop, Fava, E. (ed.), July 15-17, 1994, Center for Cognitive Science, State University of New York at Buffalo, Nemo, Padova 1995, p. 89-107.

CRESTI, Emanuela. *Corpus di Italiano parlato*. Florence, Accademia della Crusca, 2000.
CRESTI, Emanuela. Syntactic properties of spontaneous speech in the Language into Act Theory: Data on Italian complements and relative clauses. In: RASO, T; MELLO, H. (Eds.). *Spoken Corpora and Linguistic Studies*. Amsterdam/Philadelphia, John Benjamins Publishing Company, 2014, p. 365-410.
CRESTI, Emanuela; MONEGLIA, Massimo (eds). C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages [Studies in Corpus Linguistics 15]. Amsterdam/Philadelphia, John Benjamins, 2005.
CRESTI, Emanuela; MONEGLIA, Massimo. Informational Patterning Theory and the corpus-based description of spoken language. The compositionality issue in the topic-comment pattern. In: MONEGLIA, Massimo; PANUNZI, Alessandro (Eds.). *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*. Florence, Firenze University Press, 2010, p. 13-46.
FIRENZUOLI, Valentina. *Le forme intonative di valore illocutivo dell'Italiano parlato*: Analisi sperimentale di un corpus di parlato spontaneo (LABLITA), (PhD dissertation) University of Florence, Florence, 2003.
HALLIDAY, Michael A. K. *System and Function in Language*: Selected Papers. London, Oxford University Press, 1976.
't HART, Johan; COLLIER, Rene; COHEN, Antonie. *A Perceptual Study on Intonation*. An Experimental Approach to Speech Melody. Cambridge, Cambridge University Press, 1990.
HOCKETT, Charles F. A. *Course in Modern Linguistics*. New York, The Macmillan Company, 1958.
IFANTIDOU, Elly. *Evidentials and relevance*. Amsterdam/Philadelphia, John Benjamins, 2001.
FROSALI, Fabrizio. 2006. Il lessico degli ausili dialogici. In: Prospettive nello studio del lessico italiano. *Atti del IX Congresso SILFI*, Emanuela Cresti (ed.), p. 417-424, Florence, Firenze
University Press, 2006.
JACKENDOFF, Ray. *Semantic Interpretation in Generative Grammar*. Cambridge, The
MIT Press, 1972.
KALTENBÖCK, Gunther. Pragmatic functions of parenthetical I think. In: KALTENBÖCK, Gunther; MIHATSCH, Wiltrud; SCHNEIDER, Stefan (eds.). *New approaches to hedging*. Bingley, Emerald, 2010, p. 237-266.
KRIFKA, Manfred; MUSAN, Renate (eds). *The Expression of Information Structure*. Berlin,
De Gruyter, 2012.
LAMBRECHT, Knud. *Information Structure and Sentence Form*. Cambridge, Cambridge University Press, 1994.
MARTIN, P. WinPitch Corpus: a text to speech analysis and alignment tool. In: CRESTI, Emanuela; MONEGLIA, Massimo (Eds.). *C-ORAL-ROM*: integrated reference corpora for spoken Romance languages. Amsterdam/Philadelphia, John Benjamins, 2005, p. 40-51.
MAZELAND, Harrie. Parenthetical sequences. *Journal of Pragmatics*, v. 39, 2007, p. 1816-1869.
MELLO, Heliana; PANUNZI, Alessandro; RASO, Tommaso (Eds.). *Pragmatics and Prosody*. Illocution, Modality, Attitude, Information Patterning and Speech Annotation. Florence, Firenze University Press, 2011.
MELLO, Heliana; RASO, Tommaso; MITTMANN, Maryualê M.; VALE, Heloisa P.; CÔRTES, Priscila O. Transcrição e segmentação prosódica do corpus C-ORAL-BRASIL: Critérios de implementação e validação. In: *C-ORAL – Brasil I: Corpus de referência do português brasileiro falado informal*. TOMMASO, Raso; MELLO, Heliana (eds), Belo Horizonte, Editora UFMG, 2012, p. 125–176.
MONEGLIA, Massimo. The C-ORAL-ROM resource. In: CRESTI, Emanuela; MONEGLIA, Massimo (eds). *C-ORAL-ROM*. Integrated Reference Corpora for Spoken Romance Languages [Studies in Corpus Linguistics 15]. Philadelphia/Amsterdam, John Benjamins, 2005, p. 1-70.
MONEGLIA, Massimo; CRESTI, Emanuela. C-ORAL-ROM: Prosodic boundaries for spontaneous speech analysis. In: KAWAGUCHI, Y.; ZAIMA, S.; TAKAGAKI, T. (Ed.). *Spoken Language Corpus and Linguistics Informatics*. Amsterdam/Philadelphia, John Benjamins, 2006, p. 89-114.
MONEGLIA, Massimo. Spoken corpora and pragmatics. *Revista Brasileira de Linguística Aplicada*, 2011, volume 11, no. 2, p. 479–519.
MONEGLIA, Massimo; RASO, Tommaso. Notes on Language into Act Theory (L-Act). In: RASO, T; MELLO, H. (Eds.). *Spoken Corpora and Linguistic Studies*. Amsterdam/Philadelphia, John Benjamins Publishing Company, 2014, p. 468-494.
MONEGLIA, Massimo; FABBRI, Marco; QUAZZA, Silvia; PANIZZA, Andrea; DANIELI, Morena; GARRIDO, Juan Maria; SWERTS, Marc. Evaluation of consensus on the annotation of terminal and non-terminal prosodic breaks in the C-ORAL-ROM corpus. In: CRESTI, Emanuela; MONEGLIA, Massimo (eds). *C-ORAL-ROM*. Integrated Reference Corpora for Spoken Romance Languages [Studies in Corpus Linguistics 15]. Philadelphia/Amsterdam, John Benjamins, 2005, p. 257-276.
MONEGLIA, Massimo; RASO, Tommaso; MITTMANN, Maryualê M.; MELLO, Heliana. Challenging the perceptual relevance of prosodic breaks in multilingual spontaneous speech corpora: C-ORAL-BRASIL / C-

ORAL-ROM in Speech Prosody 2010. In: *Speech Prosody 2010*. Satellite workshop on Prosodic Prominence: Perceptual, Automatic Identification, Chicago, 2010. Disponível em http://aune.lpl.univ-aix.fr/~sprosig/sp2010/papers/102010.pdf.
PANUNZI, Alessandro; MITTMANN, Maryualê M. M. The IPIC resource and a cross-linguistic analysis of information structure in Italian and Brazilian Portuguese. In: RASO, T.; MELLO, H. (Ed.). *Spoken Corpora and Linguistic Studies*. Amsterdam/Philadelphia, John Benjamins, 2014, p. 129-150.
RASO, Tommaso; MELLO, Heliana. *C-ORAL-BRASIL*: corpus de referência do português brasileiro falado informal. Belo Horizonte, UFMG, 2012.
RASO, Tommaso; MITTMANN, Maryualê M. Validação estatística dos critérios de segmentação da fala espontânea no corpus C-ORAL-BRASIL. Revista de Estudos da Linguagem, Belo Horizonte, volume 17, no. 2, 2009, Disponível em http://www.periodicos.letras.ufmg.br/index.php/relin/article/view/2532/0
SCHNEIDER, Stefan. *Reduced parenthetical clauses as mitigators*. A corpus study of spoken French, Italian and Spanish. Amsterdam/Philadelphia, John Benjamins, 2007.
SCHNEIDER, Stefan. Parenthetical hedged performatives. In: KALTENBÖCK, Gunther; MIHATSCH, Wiltrud; SCHNEIDER, Stefan (eds.). *New approaches to hedging*. Bingley, Emerald, 2010, p. 267–287.
SCHNEIDER, Stefan. Les atténuateurs parenthétiques dans les textes de la Nouvelle-France du 17e et du 18e siècle. In: HELD, Gudrun; HELFRICH, Uta (eds.). *Cortesia – Politesse – Cortesía*. La cortesia verbale nella prospettiva romanistica. La politesse verbale dans une perspective romaniste. La cortesía verbal desde la perspectiva romanística. Aspetti teorici e applicazioni/Aspects théoriques et applications/Aspectos teóricos y aplicaciones. Frankfurt am Main, Lang, 2011, p. 229–252.
SCHNEIDER, Stefan. Parenthesis: fundamental features, meanings, discourse functions and ellipsis. In: MARLIES K.; OTT, D.; DE VRIES, M. (eds.). *Parenthesis and ellipsis*: cross-linguistic and theoretical perspectives. Berlin, De Gruyer Mouton, 2014, p. 277-300.
TAGLICHT, Josef. *Message and emphasis*: on focus and scope in English. London, Longman, 1984.
TUCCI, Ida. Obiter dictum: La funzione informativa delle unità parentetiche. *La Comunicazione parlata 3*, Atti del GSCP, Massimo Pettorino, Antolella Giannini & Francesca M. Dovetto (eds), vol. 3, p. 635–654, Napoli, Università l'Orientale Press, 2010.
XU, Yi. ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. In: *Proceedings of Tools and Resources for the Analysis of Speech Prosody* (TRASP 2013), Aix-en-Provence, France, 2013, p. 7-10.
ZIV, Yael. This, I believe, is a processing instruction: discourse linking via parentheticals. In: FALK, Yehuda N. (ed.). *Proceedings of the Israel Association for Theoretical Linguistics 18*, Bar Ilan University, Jerusalem, The Israel Association for Theoretical Linguistics, 2002. Disponível em http://linguistics.huji.ac.il/IATL/18/Ziv.pdf.

[RECEBIDO: agosto/2018]
[ACEITO: novembro/2018]