Modeling the prosodic forms of Discourse Markers

Tommaso Raso, Albert Rilliard, Saulo Mendes Santos

DOI: 10.14393/DL52-v16n4a2022-8

Modeling the prosodic forms of Discourse Markers

Para uma modelagem das formas prosódicas dos Marcadores Discursivos

Tommaso RASO[*]
Albert RILLIARD[**]
Saulo MENDES SANTOS[***]

ABSTRACT: This paper has a twofold goal: (i) to propose how and why to identify Discourse Markers (DM), showing that the formal features marking this category are of prosodic nature and can distinguish the six different functions of interactional nature performed by DMs. We describe both the prosodic characteristics responsible for a DM identification and the prosodic forms that convey each type of communicative function inside the more general category of DM; (ii) to show in detail the methodological steps adopted so far to allow the automatic extraction of different DMs from new data. The methodology is presented together with a statistical-computational discussion and explanation.

KEYWORDS: Discourse Markers. Prosody. Spontaneous Speech. Modeling.

RESUMO: Este artigo tem um objetivo duplo: (i) avançar uma proposta para a identificação da categoria de Marcador Discursivo (MD), mostrando que as marcas formais do MD são de natureza prosódica e também capazes de distinguir cerca de seis diferentes funções de natureza interacional veiculadas pelos MDs. Se descrevem tanto as características prosódicas responsáveis para a identificação de um MD quanto as formas prosódicas que veiculam cada tipo de função comunicativa dentro da categoria maior de MD; (ii) mostrar detalhes da metodologia que em maior medida será adotada para modelizar essas unidades e permitir uma extração automática a partir de novos dados. Ela é apresentada com uma reflexão estatístico computacional que a justifica.

PALAVRAS-CHAVE: Marcador Discursivo. Prosódia. Fala Espontânea. Modelização.

[*] Doutor em Linguística, Universidade Federal de Minas Gerais. ORCID: https://orcid.org/0000-0002-3446-313X. tommaso.raso@gmail.com

[**] Doutor em Ciência Cognitivas, Université Paris Saclay, CNRS. LISN. ORCID: https://orcid.org/0000-0001-6490-2386. albert.rilliard@limsi.fr

[***] Mestre em Linguística, Universidade Federal de Minas Gerais – Université Paris Saclay. ORCID: https://orcid.org/0000-0002-9399-9241. saulo.mendes@gmail.com

Within the Language into Act Theory (L-AcT; CRESTI, 2000; MONEGLIA; RASO, 2014; CAVALCANTE, 2020), Discourse Markers (DM) are considered interactional information units marked by specific prosodic devices with high flexibility as for their lexical fulfillment. The proposal of this paper stems from a systematic analysis of data from spontaneous speech corpora of Italian (CRESTI; MONEGLIA, 2005) and Brazilian Portuguese (BP) (RASO; MELLO, 2012), integrated by data from comparable corpora of American English (AE) and Spanish. This paper will explain this hypothesis and the methodology for its modeling.

1 L-AcT and the analysis of spontaneous speech

L-AcT is a corpus-driven theory that expands Austin's (1962) speech act theory, considering the illocution as the nucleus of the reference unit for speech and its informational patterns. The reference unit for speech is defined as the minimal stretch of speech which presents both pragmatical and prosodic autonomy. We can have two kinds of reference units: the utterance and the stanza. The utterance is made up of an information pattern around an illocutionary nucleus; the stanza (CRESTI, 2010) is made up of more subpatterns, each one with an illocutionary nucleus, juxtaposed by a prosodic continuity signal.

Therefore, the reference unit must end with a prosodic terminal boundary and is compound by one or more patterns. Each one performs an illocution and, many times, optional non-illocutionary units. Each information unit is enveloped in a prosodic unit, separated by the following one through a non-terminal prosodic boundary.

Example 1 and audio 1 show a sequence of simple utterances built up only by the illocutionary unit. Example 2 and audio 2 show a compound utterance built up by several intonation/information units, among which only the last one is the

illocutionary nucleus. Listening to audio 2a, one can perceive that the illocution is interpretable in isolation (the illocutionary unit does not need to be the last one); listening to audio 2b, one can perceive that the rest of the utterance, without the illocution, cannot be interpreted in isolation. Example 3 shows a stanza with two subpatterns, each one with its illocution (in bold). The first subpattern is linked to the following subpattern by a prosodic continuity signal (on the word *cemitério*). Both subpatterns, especially the first one, feature other non-illocutionary units that build a pattern with the illocution of their specific subpattern. Stanzas can easily present a more significant number of subpatterns.

Example 1
bpubdl03[118-128)] (áudio 1)
*GUI: faz força // mais // beleza // contrai o abdômen // joga o tronco só um pouquinho pra frente // aí // beleza // descansou // vou baixar um pouquinho mais // vai //
stronger // more // great // contract your abdomen // push the trunk just a little bit forward // ok // great // did you rest // I'm gonna slower it a little bit // go //
The utterances in example 1 are performed by a personal trainer to his client.

Example 2
bfamdl02[101] (áudio 2)
*BAO: porque / se eu for empregado / por exemplo / alguém vê que eu sou muito foda / medo de perder / o posto deles / es vão [/2] es vão me dizar né //
because / if I am hired / for instance / they see I am very good / fear of losing / their job / they will sack me //

Example 3
bfammn03[28] (áudio 3)
*ALO: aí / determinada hora lá / tava na hora de sair o [/1] o [/1] o velório / de ir po cemitério / o filho [/2] o filho mais velho vai lá dento / porque a dona Elvira até então nũ tinha aparecido cá na [/2] cá fora /=COM= né //
so / at a certain moment / it was time the wake to go out / to reach the cemetery /

the older son goes inside / become madame Elvira so far hadn't shown up outside here / did you get this //

The information units are of two types: textual and dialogic. The textual units build the semantic content of the utterance. Besides the illocutionary unit, called Comment, the textual units are Topic, Appendix of Comment, and Appendix of Topic, Parenthetical, and Locutive Introducer. Their analysis is not within the scope of this work. Dialogic units correspond to what in other frameworks are called Discourse Markers (or Pragmatic markers, Discourse particles, or other similar names).

In the last 30 years, a growing amount of literature has been published about DMs[1]. Nevertheless, it is still not clear what a DM is. More precisely, we do not have a clear answer for the following questions:

(a) How can we predict when a lexical item is a DM?

(b) Once we conclude that a lexical item (or a small sequence) behaves as a DM, how do we identify its specific function?

In our opinion, there is a fundamental problem (with few and partial exceptions) that biases the studies on DMs: the point of departure is usually the lexicon. This is very evident from the title of most of the works on DMs. The following made-up titles could exemplify the reality very closely: *The DM* well *in American conversation*; *El marcador* o sea *en el discurso académico*; *Il segnale discorsivo* cioè *nel parlato giovanile*; etc. We will try to answer questions (a) and (b) arguing that what defines both a DM and its specific function is prosody. Then, we will discuss a methodology

[1] See, among others, Bolden (2015), Degand (2014), Aijmer (2013), Traugott (2012), Fischer (2006a; 2006b), Romero Trillo (2006), Frank-Job (2006), Aijmer & Simon-Vandenbergen (2006), Schourup (1999), Brinton (1996), Bazzanella (1990), Schiffrin (1987).

for modeling different DMs according to their prosodic form, as the main feature responsible for their particular function.

2 Discussing some crucial characteristics of DMs according to the literature

Usually, in the literature, some properties are considered typical of DMs. We will discuss some of them:

(i) DMs are lexical units, or small locutions, that do not play a role at the semantic and syntactic level of the utterance since they do not partake in its propositional content; therefore, they are non-compositional items;

(ii) DMs are lexical units, or small locutions, that partially or totally lost their semantic meaning, acquiring a pragmatic function;

(iii) DMs are polyfunctional; this statement may be used in two different senses: in one sense, it means that one DM may cover different functions at the same time in the same specific occurrence; in another sense, it means that a specific lexeme may cover different functions in different occurrences;

(iv) among scholars, different functions are reported. We can divide them into cohesive and interactional functions. Cohesive functions are outside the scope of this work. As for interactional functions, the literature mentions modal functions, illocutive functions, conative function, turn-taking function, and often a strong role in politeness functions (but other functions are mentioned too).

First of all, we can say that modal functions, if we should take it as a lexeme that modifies the modality of the utterance (in the sense of BALLY, 1950), are in contradiction with the fact that there is no compositionality between the DM and the utterance. In fact, something that works as a modal operator needs to be compositional

to its semantic scope. In our view, this should be enough to say that DM are not involved in any kind of modalization. Secondly, we agree with (i) and think that DMs do not partake in the propositional content of the utterance and therefore are non-compositional with the rest of the utterance. But what marks the fact that DMs are not compositional? What makes the difference between:

(4a) *God save the queen!*
and
(4b) *God, save the queen!* (Where God is an exclamation and the rest of the utterance performs an order)

In (4b), functionally, *God* could be substituted by *Jesus* or by semantically very different exclamations, even imprecations. The function would not be affected. The choice of a specific lexeme for this function could depend on diastratic (age, cultural level, gender, specific group) or diaphasic reasons (formal or informal context, specific situation – a party with friends, a soldier of the monarchy that sees the queen is in danger, someone who is watching a chess game, etc.). In any case, we know that in (4a) *God* is the subject of *save*, while in (4b) there is no compositionality between *God* and the rest of the utterance. In (4a), we see the same thing we can also find in

(4c) See *the house of my friend!*
while in (4b) we have the same phenomenon of
(4d) See, *the house of my friend!*

Also, *see* in (4d) could be substituted by many different lexemes without losing its pragmatic function.

But the main point that these examples bring out is that the lexical form cannot be responsible for marking the loss of compositionality. How can we distinguish

between the compositional and the non-compositional status of the same lexical sequences? We can look for an answer only in prosody. The interruption of compositionality is marked by a prosodic non-terminal boundary (that is frequently transcribed with a comma). We will observe better how it is performed in spontaneous speech. The prosodic boundary may or may not coincide with a silent pause, but most of the time, in spontaneous speech, it is marked by prosodic cues other than pause.

We also agree that DMs are polyfunctional, but only if this means that a given lexeme, due to the loss of the semantic content, may perform different functions in different occurrences. In this case, what marks the function since the lexeme is the same? Again, the answer can be encountered in prosody. In fact, what marks the function is not the choice of the specific lexeme but the prosodic form of its concrete realization. Many lexemes can fulfill each function, but what conveys the functional attribution is the specific prosodic form with which the item is performed. Of course, it does exist some correlation between lexicon and function, but (i) any lexeme can perform more than one function, and in each function, it shows different prosodic characteristics; (ii) it is well known that the lexicon changes a lot diachronically, diatopically, diastratically and diaphasically. The importance of the lexicon is, in fact, much greater if we look at the effects for politeness. But this is not the main functional aspect of the DM concerning other linguistic categories. Politeness is involved in all linguistic choices, from the kind of illocution we strategically chose for our goals to any lexical choice, as well to the prosodic contour chosen to express a certain attitude. Politeness is a social category. We cannot say the same thing for the specific prosodic cues that convey grammatical functions; and more importantly, (iii) if what marks the function of a DM should be found in the lexicon, how many functions would we have, and how would we group different lexemes in the same function? We will show that, by paying attention to prosody, we can find probably six forms with interactional

functions and one form with a cohesive function. We will not discuss the latter. We refer the reader to Cresti & Moneglia (2019) for a discussion on this unit.

We still need to discuss an important aspect that we mentioned in (iv). DMs should not be confused with illocutions. Illocutions are textual units; they build the text of the utterance. This is clear in most cases since the illocutionary units are formed by more words, often many, that together build the semantic content of the most important part of the utterance. This might be not so clear in the case of illocutions expressed by exclamations, as we will see. But even in this case, since they perform an action, they can be pragmatically interpreted in isolation. They can even be the only item of a whole turn. On the contrary, DMs can never be interpreted in isolation and always depend on the illocution expressed by a different intonation unit in the utterance.

Examples (5-7) show respectively the same lexeme in a compositional context, with an illocutionary value (this means that the whole utterance is built up by just this lexeme) and as a DM. The examples show this difference using Italian, BP, and AE.

The three examples (5) show the lexemes *vedi*, *não,* and *well* in a compositional context.

> Example 5a: [audio 5a] [audio 5a1] ifamcv15[40]
> SAB*: poi / in piedi / hai visto / anche se il palco è un po' rialzato / però / se ti viene uno davanti alto / non vedi nulla //
> *then / standing up / you see / even if the stage is a little bit raised / however / if someone tall is in front of you / you don't see anything //*
>
> Example 5b: [audio es 5b] [audio es 5b1] bpubdl01[116]
> PAU*: ah / não acaba não / acaba //
> *ah / it does not end / does it //*
>
> Example 5c: [audio es 5c] [audio es 5c1] afamcv03[179]
> TOC*: &he / I didn't do terribly well there //

The three examples in (6) show the same three lexemes with an illocutionary function. In 6b, it is possible to observe that the compositional use of *não* would lead to the opposite meaning; in 6c, the lexeme is the whole content of the speaker's turn.

Example 6a: [audio es_6a] [audio es_6a1] ifamcv15 [42-45]
FER*: vedi // la metti dentro / fa finta di pigliarla / e poi la ributta fuori // vedi // non la vuole //
see // you put it inside / it seems it takes it / and then it throws it out again // see // it doesn't want it

Example 6b: [audio 6b] [audio 6b1] bpubdl01[14-15]
PAU*: não // tá dando a altura daquele que a Isa marcou lá / né //
no // it is the same height of that one that Isa marked there / isn't it //

Example 6c: [audio es_6c] [audio es_6c1] afamdl02[33-35]
*PAM: and where do you get those thoughts //
*DAR: processing what goes on around me //
*PAM: well //

Finally, the three examples in (7) show the same lexemes as DMs. The reader can listen to the audios and verify that the lexemes, which are evidently not compositional, cannot be interpreted in isolation.

Example 7a: [audio es_7a] [audio es_7a1] ifamdl02[611]
LID*: no / poi / vedi / succede questo //
no / then / see / that's what happens //

Example 7b: [audio es_7b] [audio es_7b1] bpubdl01[194]
PAU*: ah / não / ela disse que é pa ficar / por algum tempo //
ah / no / she said it must stay / for some more time //

Example 7c: [audio es_7c] [audio es_7c1] afamcv02[214]
SHR*: well / &e / you've really become &f / good friends with //

Examples (5-7) show how it is possible to answer question (a). DMs are lexemes or small locutions isolated in an intonation unit but non-interpretable in isolation. Their non-interpretability and their non-compositionality are both due to prosodic features.

3 Methodological aspects related to the prosodic modeling

3.1. Methodological premise

Throughout this work, we refer to the prosodic characteristics of DMs. For the sake of interpretability, we do not provide acoustic measurement (the statistical description and the modeling will be the object of a future work). Instead, we refer to high or low intensity, high or low f0, and long or short duration. Since spontaneous spoken data deal with different speakers and different articulation rates or intensities even in the speech of the same speaker, we need to establish a term of comparison with respect to which the prosodic parameters can be considered high, low, long, or short.

This term of comparison must be found within the utterance embedding the DM we want to describe, given the high variability of the information structure and the syntactic and lexical composition of different utterances, and given other factors that influence the speaker voice in different moments of an interaction. The only possible term of comparison is the illocutionary unit, the *Comment*. This is due to two reasons: the first one is that Comment is the only mandatory unit in order to perform an utterance (or a subpattern of the *stanza*); therefore, it is the only term of comparison we can always find. But there is also a theoretical reason: in the L-AcT framework, the illocution is the nucleus of the utterance, while all the other units build a pattern with the Comment and are informationally and prosodically subordinated to it.

Therefore, whenever we say that a DM is marked by a high or low intensity, high or low f0, and long or short duration, we mean that it has these characteristics

with respect to the syllabic mean of the Comment. Of course, the Comment can express different illocutions and therefore can have different prosodic forms; however, its nucleus (the chunk responsible for prosodically marking the illocutionary value in a Comment unit) is composed of a few syllables, normally just one or two, and this strongly reduces the impact of this variability on the mean syllabic measurements.

3.2 Acoustical analysis of prosodic correlates

The voice fundamental frequency (F0), its intensity, and the rhythm of enunciation form the most important correlates of the cognitive processing of the prosody main functions (HIRST; Di CRISTO, 1998). Their estimation and interpretation from the recording are the subjects of a large bibliography that shows the complexity of the task (for a review, see, e.g., COLE, 2015). If pitch estimation is now routinely proposed in many software suites (PRAAT, WINPITCH, SpTK etc.), the output is not systematically reliable. Particularly, non-modal phonations (GERRATT; KREIMAN, 2001) raise significant difficulties regarding the evaluation of a frequency (i.e., a regular phenomenon), whose definition is problematic when vocal folds vibration mechanisms are anything but regular. In Brazilian Portuguese, notably, post-stress syllables generally receive much lower energy, which tends to lead to complete or partial devoicing, or to the apparition of non-modal mechanisms. In English, many studies have shown utterance-final creak functions as a boundary marker (e.g., DAVIDSON, 2019). Martin (2012) shows a sample of detection errors that links particular deteriorations of the speech signal quality to various types of pitch detection algorithms: each pitch detection algorithm (PDA) appears to show specific robustness and weaknesses to different alterations of the input signal. To that aim, it may be interesting to compare the results of several PDAs on a given target signal so as to be able to detect potential problems and zones of high reliability of the F0 detection. The

current ongoing work proposed, for example, an estimation of F0 on the target corpus using the following PDA: Praat's ac, cc, and shs algorithms (BOERSMA; WEENINK, 2020), yin (DE CHEVEIGNÉ; KAWAHARA, 2002), swipe (CAMACHO; HARRIS, 2008), rapt (TALKIN, 1995), and openSMILE:) (EYBEN; SCHÜLLER, 2015). Once candidates from each algorithm are obtained, the output F0 vectors are time-aligned so as to deal with variation in the sampling frequencies and time alignment of each PDA, using typically a linear interpolation. Interpolated F0 candidates are then compared between PDA so as to detect if and where some may show differences in the estimated F0: a "gross error rate" variation above 20% between two candidates is generally considered problematic, as most errors are linked to octave jump (SIGNOL; BARRAS; LIENARD, 2008; CAMACHO; HARRIS, 2008). Note that in our case, unlike the situation in a typical evaluation process, there is no ground truth or reference value: all the candidate PDA algorithms potentially output reliable or erroneous estimations. In a first approximation, the value which is agreed by the majority of the tested algorithms was selected here; in the future, a dynamic programming algorithm will be set up so as to select the candidate (among the proposals of all algorithms) that satisfies criteria of continuity (smooth curve) and majority decision (see, e.g., VINCENT; ROSEC; CHONAVEL, 2006, for an example of application).

Of a different nature than F0, the signal's intensity is relatively straightforward to estimate, but its uses as a reliable correlate of perceived loudness are also problematic: first, its absolute level (sound pressure level) is generally lost during the recording procedure due to many factors (for details and solutions, see ŠVEC; GRANQVIST, 2018). For the intensity value to reflect the perceived loudness, it shall also be submitted to a specific weighting of its frequencies: typically, the A-weighting was designed to reflect human ear perception. Loudness also has complex relations with pitch and the spectral characteristics of the sound (MEUNIER et al., 2018), thus one may use dedicated models of loudness to express the perceived strength of speech

sounds, despite the complexity of such tools (MOORE et al., 2016). Another avenue to estimate information linked to the energy of the signal is to estimate changes in the signal linked to vocal effort (TITZE; SUNDBERG, 1992; LIÉNARD; DI BENEDETTO, 1999; TRAUNMÜLLER; ERIKSSON, 2000); this is basically done via estimations of the energy decrease in the spectrum since an important effect of higher effort is to raise the spectral slope (NORDENBERG; SUNDBERG, 2004; SUNDBERG; NORDENBERG, 2006). Meanwhile, vowel articulation has a major impact on the spectral slope when estimated on the speech signal: the proposed measurements of spectral emphasis thus generally rely on the long-term average spectrum (LTAS) with speech sequences of about 20 seconds so as to stabilize the spectrum (about 20 seconds of speech, from which about 10s. of voiced framed are extracted; see LÖFQVIST; MANDERSSON, 1987). From such LTAS, based on voiced frames only, the following indexes are often used in the literature: the so-called "Hammarberg index", which is the difference between the peak amplitudes of the 0-2 and 2-5kHz bands of the power spectrogram (HAMMARBERG et al., 1980); the alpha index, which is the ratio of sound energy above and below 1kHz (SUNDBERG; NORDENBERG, 2006), or the "spectral emphasis" as defined by Traunmüller & Eriksson (2000), which is the spectral energy above 1.5 the mean F0 on the concatenated voiced segments. These indexes are all derived ways of estimating the increase of spectral slope produced by vocal effort, typically to estimate changes linked to arousal (BANSE; SCHERER, 1996; GOUDBEEK; SCHERER, 2010). Liénard (2019) shows that it is possible from LTAS to estimate the original "voice strength" thanks to the characteristics of the low and mid-range of the spectrum, for gender-specific data, with an accuracy of 3dB. The main limitation of these measurements, with regard to an approach that is looking for short-term measurements, is the fact they need long-term speech data (as defined above) to be reliably evaluated (reliably meaning without the effect of the segmental changes produced by articulation). Another possibility would be to estimate parameters of the

source component related to loudness (see table 2 in D'ALESSANDRO, 2006, for propositions) thanks to an inverse filtering approach; unfortunately, the inverse filtering is not a reliable process on sustained vowels from many different speakers (KREIMAN; GERRATT; ANTOÑANZAS-BARROSO, 2007), hence shall give even worst results from spontaneous speech recorded during field works.

Considering the limitations of both raw intensity estimation from uncalibrated signals and the fact that reliable vocal effort estimations (in dB) can only be done on long-term signals, a potential approach may consist in estimating the mean effort of a speaker from the longest possible connected utterance at hand, containing the targeted unit from which intensity is to be extracted, and using this value to "calibrate" the mean intensity of the complete utterance. The intensity of the targeted unit would then be normalized with respect to the mean calibrated intensity of the sentence, following equation 1, where the calibrated intensity I_c estimated at time t equals the mean vocal effort E estimated (in dB) on the complete utterance plus the difference between the observed raw intensity at time i, $I_r(t)$ minus the mean of the raw intensities estimated on the N samples of the utterance.

Eq. 1: $I_c(t) = E + \left(I_r(t) - \frac{1}{N}\sum_{x=1}^{N} I_r(x)\right)$

The problem is, thus, the estimation of the mean effort, or voice strength in Liénard's terms (2019), expressed in dB: in the absence of a reliable model to estimate such value (there are none currently available to our knowledge), the mean intensity of the sentence, or of a reference part of the sentence (typically, in this context, the Comment), may be used as a reference so as to express the intensity measurement differentially regarding the reference illocutionary nucleus.

Manual speech segmentation is a relatively simple task, if heavily time-consuming. Meanwhile, deriving the perceived rhythm from the raw duration of segments is also non-trivial due to phoneme-specific intrinsic and co-intrinsic durational constraints. Normalization procedures have been developed, based on the standardization of raw durations with regard to the observed distribution of duration for similar segments in similar contexts (CAMPBELL; ISARD, 1991): this statistical approach allows an efficient estimation of segmental lengthening, expressed as the deviation from the expected duration for a phoneme with a set of characteristics. Building on this approach at the segmental level, Barbosa (2007) developed a model of speech rhythm that estimates the lengthening for syllable-like units (the so-called vowel-to-vowel unit) according to the duration characteristics of its phonemes and the contextual lengthening. This model (and other information) was then implemented into a semi-automatic tool available to the research community so as to derive reliable measures of rhythm from segmented speech (BARBOSA, 2013).

3.3 Stylization & normalization of measurements

Once reliable measurements of acoustic parameters are available, one may try to extract information from this dataset so as to describe what in the acoustic characteristic of prosody is related to the targeted functional aspects. Not all changes in the measurements are relevant to prosodic functions. For example, articulation has a dramatic effect on the measured intensity, typically with consonantal articulation. Intrinsic phonemic differences are also observed in the F0 and intensity values, similar to what was described earlier for duration. Such changes in parameters are regrouped under the umbrella term of microprosodic effects (DI CRISTO; HIRST, 1986; DUBĚDA; KELLER, 2005), where the articulation constraints produce systematic and predictable changes (e.g., higher F0 on close vowels) on both measurements. These changes may

be relatively large, but on vowels, they generally do not exceed perception thresholds; the changes are larger for co-intrinsic factors, with consonantal articulation having important effects on vowel's F0 – a fact that helps phonemic identification (HONDA, 2004).

If important for segmental perception, microprosody introduces changes that are not of interest for the macroprosodic aspects of speech and are more a nuisance to prosodic analysis. F0 is also a measurement that changes continuously throughout an utterance, while the pitch perceived on a given syllable may often be represented by a single note by trained musicians (see musical performances of Hermeto Pascoal, for example). This is partly due to the integration of the F0 value as a single note by our perceptual system for variations below a given threshold of perception (ROSSI, 1971). This threshold has a (negative) linear relation with the duration of the tone, on logarithmic scales ('T HART; COLLIER; COHEN, 1990) – i.e., the longer the sound, the smaller the threshold of perception for pitch differences. These facts lead the researchers of the IPO school to propose a straight-line stylization of the original F0 curve ('T HART et al., 1990), with line segments fitting the raw measurements, so the resynthesis of the stylized version is indistinguishable from the original (their so-called "close-copy stylization"). The stylized pitch curve has (at least) two main advantages over the raw measurements: (i) it proposes an economical description of F0 changes in terms of quantity of information required to describe the curve; (ii) it removes changes that do not participate to the perceptual processing of prosody, so changes that have no interest for prosodic models. Several variants of IPO close-copy stylization have been proposed in the literature that departs from 't Hart et al. (1990) proposal on several aspects but kept the perceptual equivalence principle. Two renowned implementations may be cited here: Hirst & Espesser (1993) MOMEL algorithm, based on quadratic spline functions producing a continuous smooth curve stylization of a sentence melody, and d'Alessandro & Mertens (1995) model of tonal perception that

stylizes F0 variation on each vocalic segments by straight lines, and is implemented as a Praat script in the Prosogram (MERTENS, 2004).

3.4 Parametric description of prosodic variations

F0 stylization is an important step in simplifying the F0 curve so as to remove some of its undesirable characteristics (microprosody, declination line), keeping the meaningful features that are thus described via an economical set of parameters. Phonological approaches to the prosodic description would take a further step after close-copy stylization, reducing even more F0 changes so as to keep only variations that are relevant to phonological functions. This led 't Hart et al. (1990) to propose their standardized stylization – a straight line curve that is no more perceptually equivalent to the original (in a psycholinguistic approach) but carries the same (phonological) functions as the original; they propose that the rising and falling movements of their stylization are the basic elements of their proposed grammar of intonation. Other implementations of phonological models have started with similar principles (close-copy stylization, based on various processes), but then proposed the target levels of the stylized pitch movements as the basic elements descriptive of the phonological models – e.g., the INSINT alphabet (HIRST; DI CRISTO; ESPESSER, 2000), Pierrehumbert (1980, 1981) High and Low targets, or the Polytonia system based on the Prosogram (MERTENS, 2014). Note that these systems of phonological description (without tackling here position on their theoretical differences, see the discussions in the referred literature for details) may be used for the latter processing, especially in cases where phonological functions are targeted, using their respective descriptive alphabets as feature set – see Rilliard (2019) for an example of application. Meanwhile, the functions that are targeted here are implemented at a pragmatic level; hence we will prefer keeping it at the phonetic level for the description of prosodic variation.

From a close-copy stylization, it is possible to extract a set of parameters that will describe the features of the prosodic variation. This may be done in several ways. 't Hart et al. (1990), for example, propose a feature matrix (rising, falling, etc.) that describes the changes in the stylized prosody, along the same lines as a proposition by Martin (1973, 1987) or Contini & Profili (1989). One may also describe the prosodic changes in terms of the fit of the estimated F0 points (raw or stylized ones) by polynomial functions; this is particularly efficient in cases where the prosodic units that are compared have a comparable duration. This can be done by fitting orthogonal polynomials to the measured F0 points and using the coefficients of these polynomials as descriptors of the underlying intonation shapes (LEVITT; RABINER, 1971; KOCHANSKI; SHIH, 2003; GRABE; KOCHANSKI; COLEMAN, 2007; LAI, 2014). Note the MOMEL approach (HIRST; ESPESSER1993) may fall in the same category, if not based on polynomials but on splines. Similar curve-fitting approaches are found with the application of the Fujisaki model (FUJISAKI, 1983, 1988) to prosodic description (MIXDORFF, 2000): the parameters of the fitting model are here motivated by the F0 production mechanisms, and they allow the description and comparison of prosodic performances (MIXDORFF; PFITZINGER, 2005; GURLEKIAN et al., 2010). Other model-based approaches have been followed, but without the motivation that characterizes Fujisaki's model, using a functional data analysis framework (FDA, RAMSAY; SILVERMAN, 2005): these works take advantage of the descriptive power of this statistical framework so as to compare the prosodic characteristics of the investigated functions (EVANS et al., 2010; HADJIPANTELIS; ASTON; EVANS, 2012; CAVALCANTE, 2020). The main point of all these approaches is their ability to describe and compare the shapes of F0 contours – thus potentially catching an important aspect of intonational variation, as in the case of tones (EVANS et al., 2010) or nuclear accents (GRABE et al., 2007).

The question of the comparability of the prosodic units, in terms of duration, is a complex one: there always is some durational variations between performances of similar speech "items"; this is easily overcome by introducing some temporal normalization – e.g., by considering a fixed number of points along the targeted contours (XU, 2005) – but such a solution is valid only if the structures of the compared objects are similar, typically in terms of the number of syllables. Not taking care of this may lead to misalignment of the different structures – while alignment of F0 with the underlying syllabic and lexical structure is a crucial prosodic characteristic (KOHLER, 2006).

Another way to describe prosodic variation and to compare its performances is to extract sets of summary statistics from each item to be characterized. This typical machine-learning approach will extract a set of low levels descriptors (e.g., F0, intensity, rhythmic measurements, several voice quality-related parameters, and often spectral information) and propose statistical descriptors of them across the complete item or relevant parts of the item (the nucleus, for example). Such statistical descriptors are typically the mean, standard deviation, first and second derivative, etc. Barbosa (2013) proposes a Praat script that allows the extraction of these kinds of descriptors; software like openSMILE (EYBEN; SCHULLER, 2015) was built so as to allow the extraction of feature-rich vectors then used for the categorization of, e.g., affective variations (EYBEN et al., 2016; MACARY et al., 2021). A frequent limitation of these approaches (for linguistic studies) is the predominance of black-box algorithms used for the classification tasks, which prevents an interpretation of the results, together with the large and mostly uninterpretable feature set. It is nonetheless possible to follow this path using smaller (and motivated) feature sets and to use white-box statistical algorithms.

3.2 Classification or clustering of observations

One aim of this methodological paper is to describe possible paths in the description of prosodic variations related to the description of communicative functions. Once the steps described in the previous parts have been done, the researcher has at her/his disposal a set of prosodic features on the one hand and communicative functions related to the items observed in a given corpus on the other; the problem being thus to describe/find the relations between these two groups.

There are two possibilities here (for details and more in-depth descriptions of possible options, the reader is referred to, e.g., VENABLES; RIPLEY, 2002; RENCHER; CHRISTENSEN, 2012): or the theoretical categories (communicative functions, or other – here the functions of the MDs) are already known for each item, or they are not. In the first case, classification algorithms, or tasks of supervised learning (in the machine learning world and speak), is the most obvious path; a classical one is the use of a discriminant analysis approach in the search for the best combination of the available prosodic features able to *predict* the theoretical categories. The aim is to build a predictive model able to attribute one of the theoretical functions (closed set) to any new observation (i.e., one not part of the training set used to build the model); a derived use is to describe how the prosodic features are combined to allow this predictive capacity, so as to better understand possible relationships between the two levels – i.e., one may use a white-box algorithm so as to be able to observe the specific feature combination build by the model, the aim not being the best possible predictive capacity, at the expense of descriptive ability. Note that any type of prosodic features (curve fitting parameters or feature vectors, or also characteristics extracted from phonological models) may be used in this process. In the second case, the theoretical categories are not previously known, and the aim is more descriptive, unsupervised, or exploratory (in the sense of VENABLES; RIPLEY, 2002): how the measured prosodic

features do separate the collected items into groups (or "clusters") that share similar prosodic characteristics within a cluster and maximal differences across clusters. These methods help observe patterns in the data that are able to show similarities within a subset and differences between subsets. The interpretative work – how the obtained groups or dimensions do relate to the theoretical communication functions – is left, of course, to the researcher; note also this process is potentially limitless, as the potential combinations are open.

One of the first steps in an exploratory process is the visualization of the dataset, generally using techniques linked to Principal Component Analysis (with or without rotations). This allows the reduction of the complexity of the prosodic features by finding the correlations between them and potentially allows observing potential patterns in the cloud of points formed by the observations. An approach of widespread use in the linguistic literature (e.g., NERBONNE et al., 2011) is the Multidimensional Scaling algorithm (MDS), which seeks a projection of the data set onto a reduced (2 or 3) number of dimensions, a dimensionality that best fit the geographic spread of dialectological data (but not necessarily the complexity of pragmatic functions). The method is based on a dissimilarity matrix that allows the comparison of a "distance" between each pair of items. The computation of such a (dis)similarity amounts to being able to calculate the prosodic "distance" between items. This prosodic difference, if one's goal is to link prosodic parameters to communicative functions, has to reflect the perceptual difference listeners would make between these functions. There is no perfect match between a set of prosodic features and their combination to propose an objective dissimilarity and the perceptual grouping obtained in different perceptual evaluation, but it is exactly this goal such techniques would ultimately try to produce (HERMES, 1998a, 1998b; HIRST; RILLIARD; AUBERGÉ, 1998; DE CASTRO MOUTINHO et al., 2011; D'ALESSANDRO et al., 2014).

4 Results: description of DMs' prosody

4.1 Raw characteristics of main DM categories

4.1.1 Previous studies

The proposal that Dialogic Information Units could explain DMs was firstly advanced by Cresti (2000), where an initial distinction was made. Cresti, studying Italian, proposed that there were four DMs (Dialogic Units in the L-AcT terminology), which she called Incipit (INP), Conative (CNT), Allocutive (ALL), and Phatic (PHA).

Incipit is described as having the function of taking the turn or beginning the utterance, expressing affective contrast with the previous utterance. Its distribution is always the initial one (in the utterance or in the stanza's subpattern), and its prosody shows very high intensity and f0, and a very short duration. Cresti says that INP can have different forms: rising, falling, and rising-falling. This poses one problem: why the same function should be conveyed by different forms? We will answer this question below.

Conative is described as having the function to push the listener to do something or to stop doing something. It has a free distribution. Prosodically, it is described with a falling profile, short duration, and high intensity. Also, Allocutive is described as distributionally free, but we will be back to this point later. Functionally, it serves to establish social cohesion among the interlocutors or to disambiguate who is the addressee of the utterance, using titles, epithets, and proper names. ALL is the DM where the lexical category is more defined. Cresti (2000) attributes to ALL a form similar to CNT: falling movement, short duration but low intensity. The description of these two units generates confusion when CNT is fulfilled by the lexicon typical for ALL, which is not rare. In fact, this has led to overestimating the number of ALL and underestimating that of CNT. As we will see below, a better prosodic description of these two DMs allows a clear distinction.

Phatic is described functionally as serving to maintain the channel open. This function seems too vague. It could be attributed to any kind of filled pause or even to textual informational units scanned in more intonation units. Besides, its prosodic description emphasizes the very short duration and very low intensity of the unit but does not individualize a specific form, which would be explained by its very low duration. Distributionally, it would be free too. This unit, therefore, poses some problems: we do not have a clear function for it, its prosodic form is not defined, nor its position could help its recognition. We will pick up on these difficulties below.

Frosali (2008), using the same theoretical framework, proposed two more DMs: Expressive (EXP) and Discourse Connector (DCT). Since DCT (CRESTI; MONEGLIA, 2019) has a cohesive function, it will remain outside of the scope of this work. As for EXP, it is described as follows: functionally, it would express emotional support for the illocution; distributionally, it would be free; and prosodically, it would present medium intensity and duration, and a form defined as "modulated", meaning that it is possible to see more than one small f0 movement in it. Evidently, also EXP poses some problems: what would it really mean to have the function of emotionally supporting the illocution? Besides, we need to have a clear prosodic form in order to say that DMs could be grouped in a limited set of prosodic cues mainly responsible for conveying their function. Also, this case will be better explored.

Raso (2014), using Cresti's and Frosali's categories and descriptions, tries a first systematization of the proposal studying two comparable sub-corpora of C-ORAL-ROM Italian and C-ORAL-BRASIL. At this point, it became clear that a better prosodic analysis was needed. This was the goal of Raso & Vieira (2016). They found an answer for the different profiles of INP and turned clear the prosodic distinction between ALL and CNT. But they leave the other DMs still in need of a better understanding.

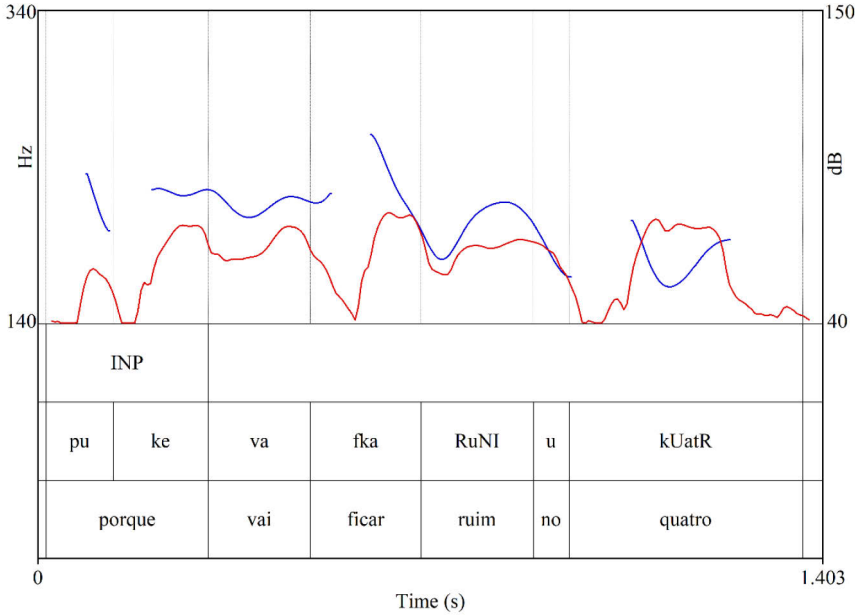
4.1.2 Incipit

As the examples below show, the different forms of INP are due to microprosodic effects. In this specific case, these effects are important because they are magnified by the very high f0 and the very short duration of this unit. Its form, in fact, should be described as a flat stressed vowel with a very high range, higher than the Comment mean, and very high intensity, higher than the Comment mean. Nevertheless, if the stressed syllable is preceded or followed by consonants or another syllable, its form is affected by the microprosodic segmental material; a non-stressed syllable before (or a single voiced consonant) produces a rising movement, in order to reach the high pitch; a non-stressed syllable (or the semivowel of a diphthong) after the stressed vowel produces a falling movement. These movements have a very high f0 variation rate.

Example 8 [audio 8] bfamcv03[257]
*CEL: porque /=INP= vai ficar ruim no quatro //
because / it will be bad for the four //

Example 8 (fig. 1) shows how the prestressed syllable is influenced by the micromelodic effect of the unvoiced bilabial, while the stressed one is influenced by the unvoiced velar, whose f0 range is maintained by the stressed vowel.

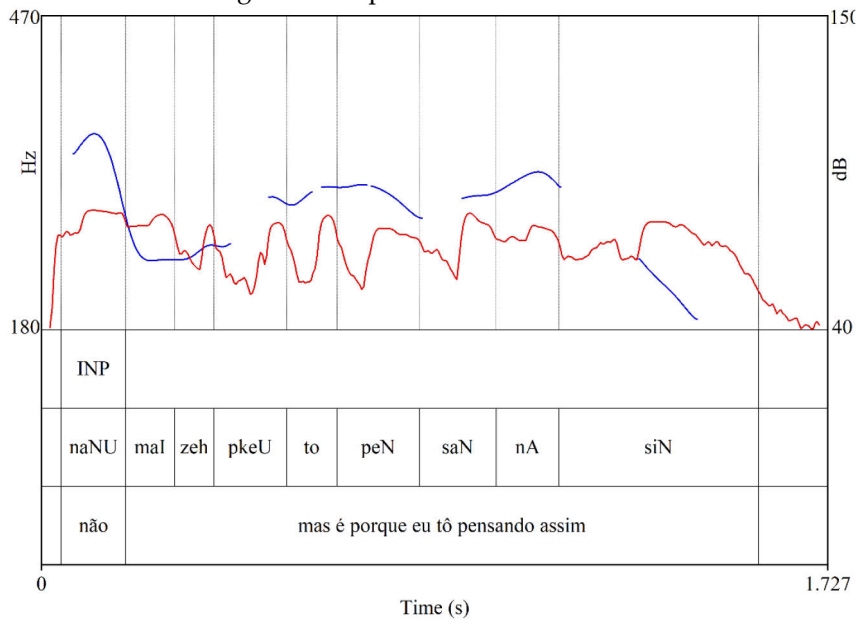
Figure 1 – Representation of audio 8: f0 (in blue), intensity (in red), tagging of the target unit, broad syllabic transcription, and orthographic transcription of the utterance[2].



Source: Raso & Vieira, 2016.

Example 9: [audio 9] bfamdl02[195]
*BAO: não /=INP= mas é porque eu tô pensando assim //
no / but it is because I'm thinking this way //

Figure 2 – Representation of audio 9.



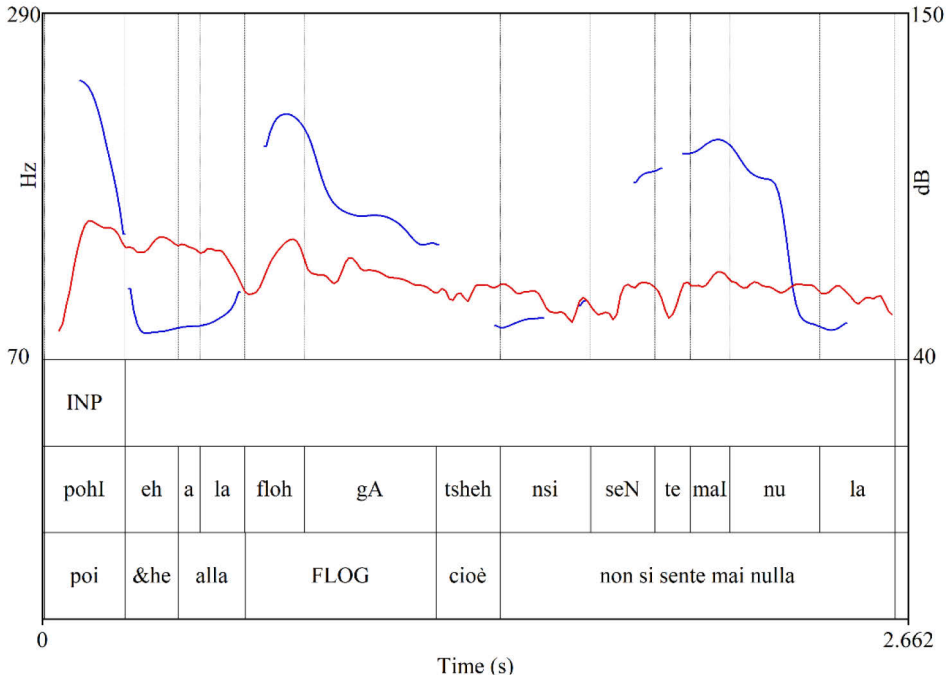
Source: Raso & Vieira, 2016.

[2] All the figures showing the different units have the same structure.

Figure 2 shows how the form of the INP is affected by an initial voiced consonant that causes a rising movement until the vowel of the diphthong, which falls on the semivowel.

Example 10: [audio 10] ifamcv06[32]
*ILA: poi /=INP= alla FLOG / cioè / non si sente mai nulla //
The / at the FLOG / I mean / we never listen to anything //

Figure 3 – Representation of audio 10.

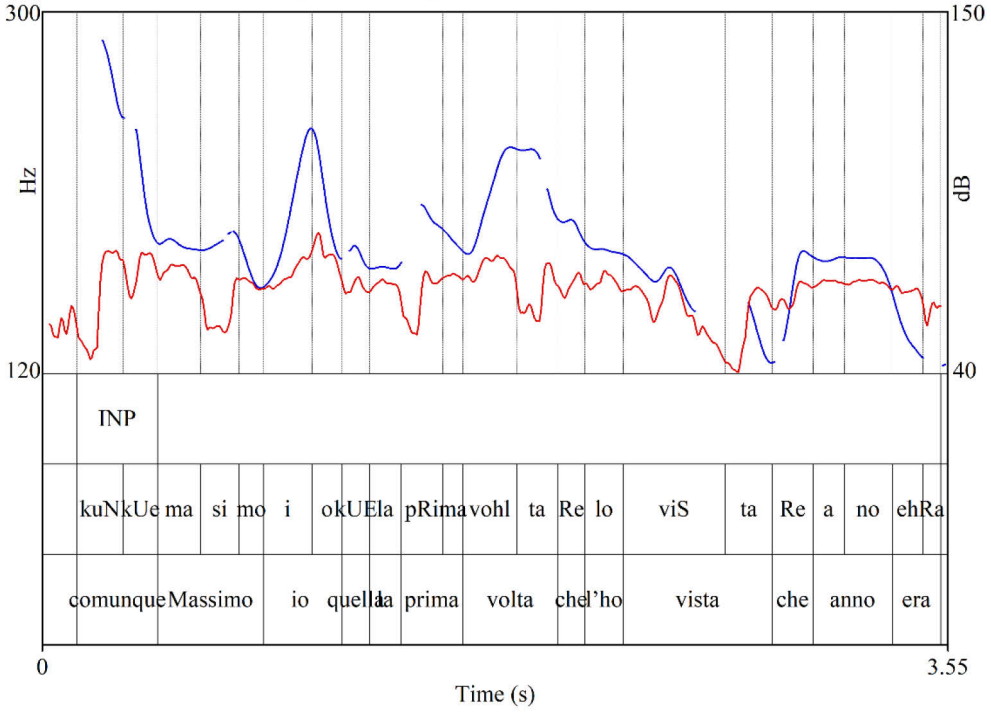


Source: Raso & Vieira, 2016.

Figure 3 shows the effect of the unvoiced consonant, while, like in figure 2, the semivowel of the diphthong causes a falling movement.

Example 11: [audio 11] ifamcv01[871]
*ELA: comunque /=INP= Massimo / io / la prima volta che l'ho vista / che anno era //
anyway / Massimo / I / the first time I saw her / what year was it //

Figure 4 – Representation of audio 11.



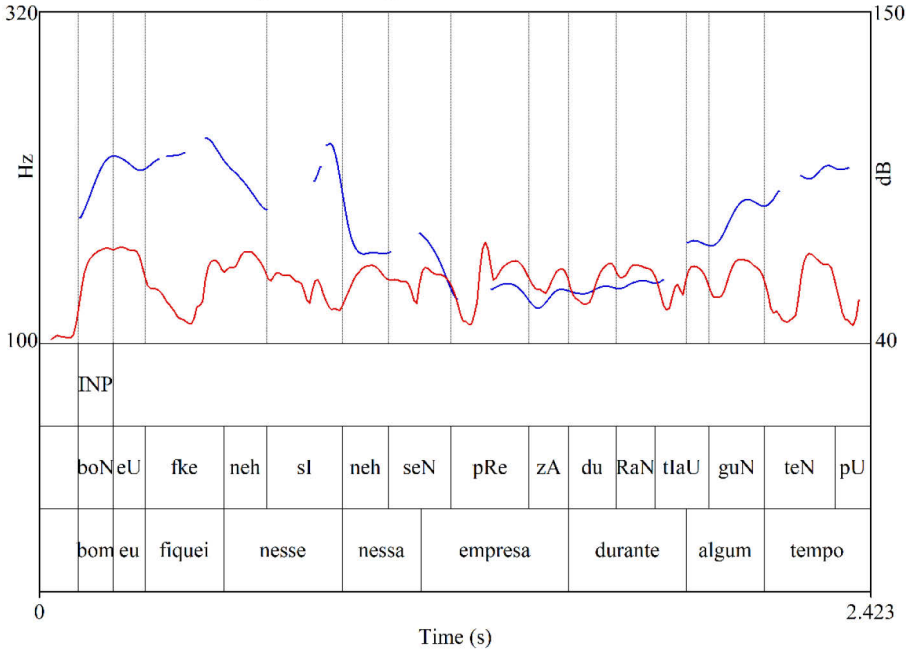
Source: Raso & Vieira, 2016.

Figure 4 shows the high range of the beginning of the INP due to an unvoiced consonant and the falling profile of the unstressed syllable. Note that the Italian word *comunque* [ko'mũkwe] is pronounced [kũkwe].

Example 12 [audio 12] bfammn06[49]
*JOR: bom / eu fiquei nesse [/1] nessa empresa durante algum tempo /
well / I remained in this firm for a while /

In figure 5, it is possible to see the rising profile due to a voiced consonant that needs to reach the high range of the stressed vowel, not followed by any segmental material.

Figure 5 – Representation of audio 13.



Source: Raso & Vieira, 2016.

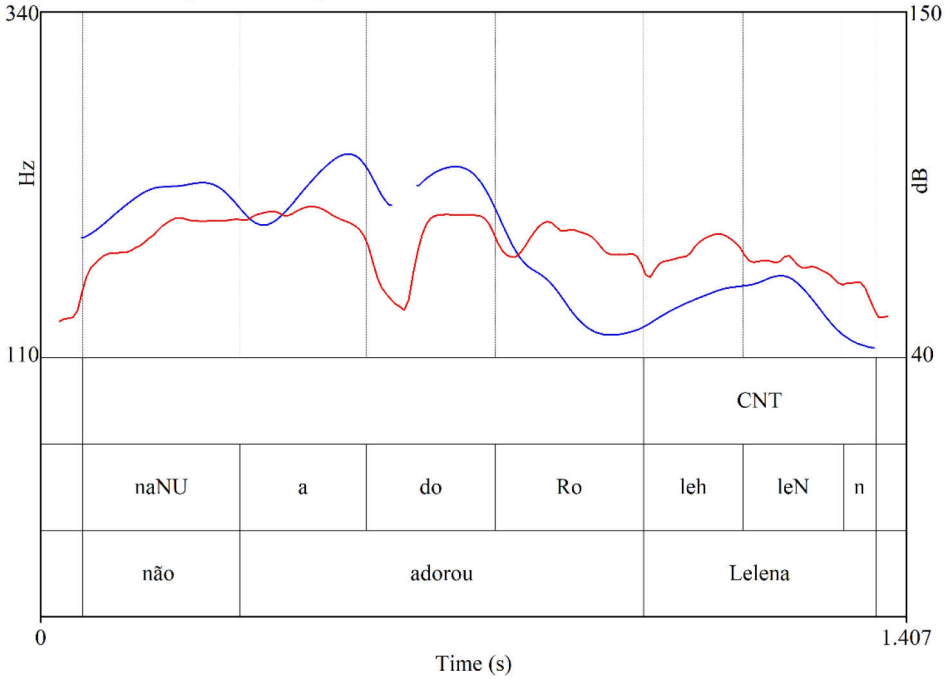
4.1.3 Conative and Allocutive

The examples below show the differences between CNT and ALL. CNT has a falling movement, usually with a high f0 variation rate (even if not so high as in INP and not higher than the Comment) and a high intensity (but not so high as in INP and not higher than the Comment). But what is more interesting is that in CNT the falling movement is aligned with stress syllable, while ALL falls since the beginning. Later (RASO; FERRARI, 2020), it was also observed that in CNT, before the falling movement, there is a slightly rising movement. This is more evident when the stressed syllable is not the first one, but it can be observed even if the stressed syllable is the first one and there is enough segmental material before the vowel; of course, it cannot be seen if this material is a non-voiced consonant. Raso & Ferrari (2020) propose a probably clearer functional definition of CNT. It would signal the illocutionary solution of the utterance.

Looking more in-depth at the behavior of ALL, Raso & Ferrari (2020) noted that it cannot appear in the initial position and prefers the final one. However, the few cases of ALL in medial positions help to better understand its form. ALL has a falling movement that becomes flat in its second part, regardless of the position of the stress. Most of the ALLs are in the final position, which partially masks their real form. In fact, due to the end of the utterance and to the fact that ALL features a low intensity, the falling part is more evident, and the flat part is frequently non-articulated or with such a low intensity that its f0 cannot be tracked.

Example 13 : [áudio 13]
*LUR: não / adorou / Lelena //=CNT=
no / he liked it // Lelena //

Figure 6 – Representation of audio 13. ifamdl12[238].

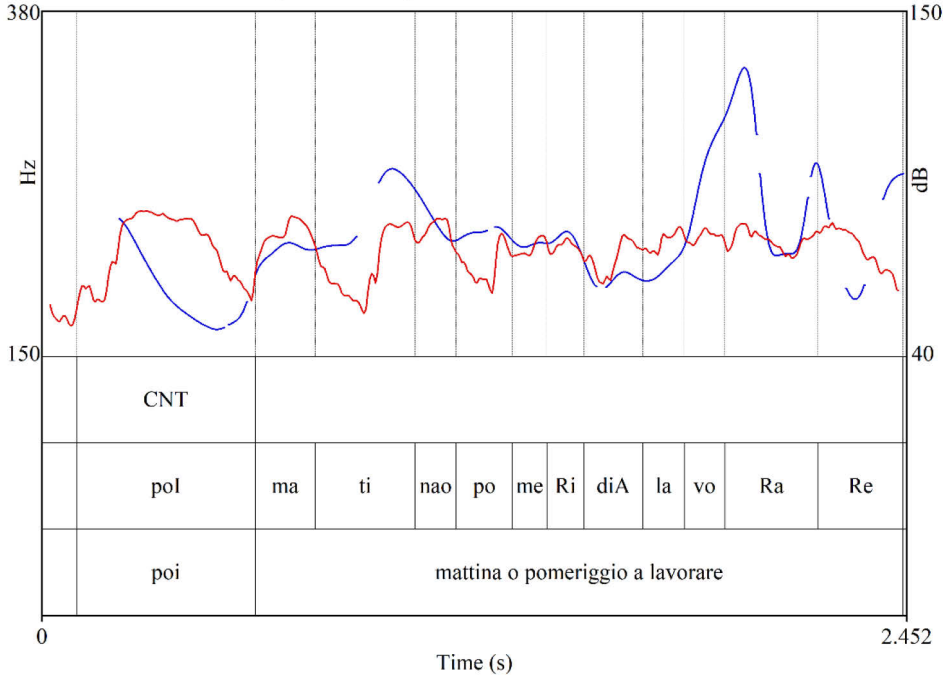


Source: Raso & Vieira, 2016.

Figure 6 allows us to observe the alignment of the falling f0 movement on the stressed syllable and the rising movement in the prestressed one.

Example 14: [audio 14] ifamdl12[235]
FRA*: poi /=CNT= vabbé / mattina o pomeriggio a lavorare /
then / okay / morning and afternoon working /

Figure 7 – Representation of audio 14.



Source: Raso & Vieira, 2016.

This figure can be compared with example 10 (fig. 3), where the same lexeme *poi* has the function (and the prosodic form) of INP. In figure 7, the CNT does not have the possibility to feature the initial rising movement since the stressed vowel is preceded by an unvoiced consonant.

The next four figures aim at showing once more how the lexicon cannot be considered a strong functional vehicle. In examples 15, 17, and 18 (fig. 8, 10, and 11), we have the same lexical item *Rena* (from the name *Renata*) working respectively as a calling illocution, as an ALL, and as a CNT. In 16 (fig. 9), we have another proper name, with the same accentual structure; also, in this last case, as in 15 (fig. 8), the name conveys an illocution, but it is not a calling; it is a warning. The prosodic form of calling in 15 (fig 8) is clearly different from that of warning in 16 (fig.9). Likewise, the prosodic

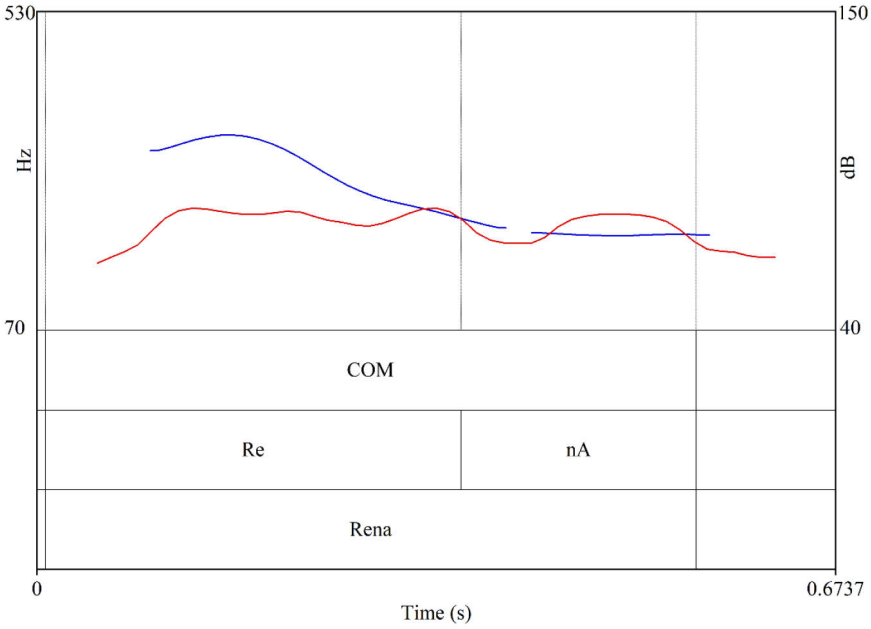
forms in 17 (fig. 10) and 18 (fig. 11) are different both with respect to each other and with respect to the two illocutionary forms.

In the calling illocution, the name *Rena* features a duration of 476 ms, an f0 mean of 248 Hz, reaching its maximum at 355 Hz, and mean intensity of 78 dB. In the warning illocution, the name Bruno features a duration of 272 ms, an f0 mean of 210 Hz, reaching its maximum at 263 Hz, and mean intensity of 84.6 dB. These parameters are much different from those of example 17 (ALL) and 18 (CNT), uttered by the same speaker of 15. In example 17 (fig. 10), Rena has a duration of 203 ms, an f0 mean of 335 Hz (due to the very high pitch of the whole utterance), and a mean intensity of 72 dB. In 18, Rena features a duration of 371 ms, a mean f0 of 232,5 Hz, with its maximum at 258 Hz, and a mean intensity of 70,5.

Some of these prosodic aspects can be evaluated only with respect to the specific illocution of the utterance, but what is interesting is to observe (i) that ALL and CNT are much shorter than the illocution of calling (as explained in note 6, the warning needed to be very fast due to a clear attitude of urgency); (ii) that, in both illocutions, a prominence with its specific illocutionary form is clearly recognizable; (iii) that, in the ALL, the f0 is falling from the beginning, but in the CNT it starts at 230 Hz, reaches 258 Hz and then falls to 204 Hz; the falling part starts three or four pulses after the beginning of the vowel.

Example 15: [audio 15] bfamdl01[255]
FLA*: Rena //=COM=

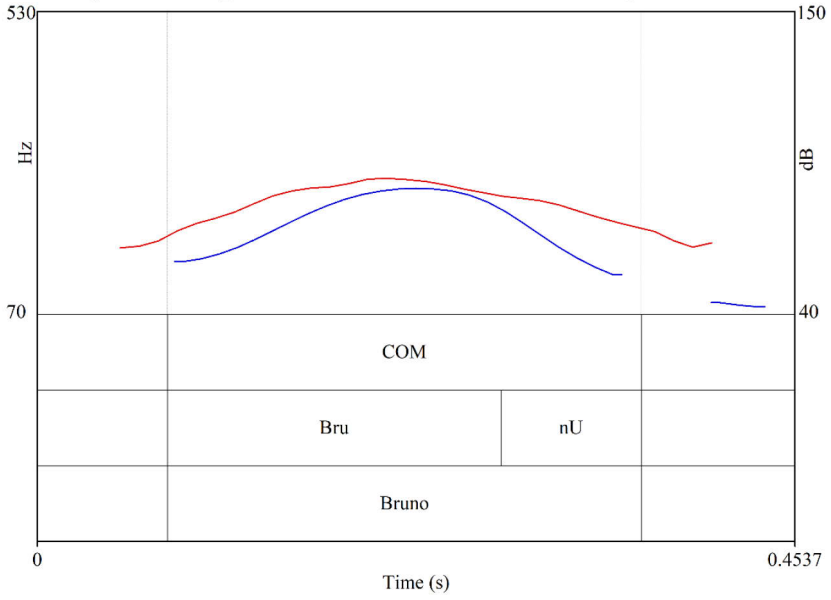
Figure 8 – Representation of audio 15 – Calling illocution.



Source: Raso & Ferrari, 2020.

Example 16[3]: [audio 16]
DEI*: Bruno //=COM=

Figure 9 – Representation of audio 16 – Warning illocution.

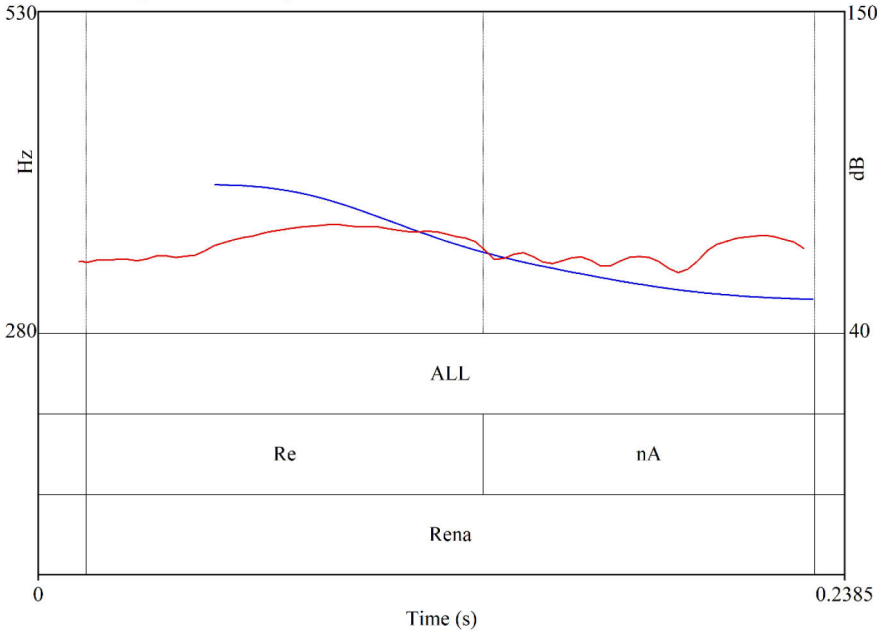


Source: Raso & Ferrari, 2020.

[3] This example is extracted from a corpus of Angolan Portuguese not yet published (ROCHA et al. 2018). In this context, *Bruno* is walking together with an Angolan guide while a truck is dangerously approximating. The "vocative" by the guide is a warning illocution, not a calling.

Example 17: [audio 17] bfamdl01[194]
FLA*: (&va [/1] vai esse / né /) Rena //=ALL=

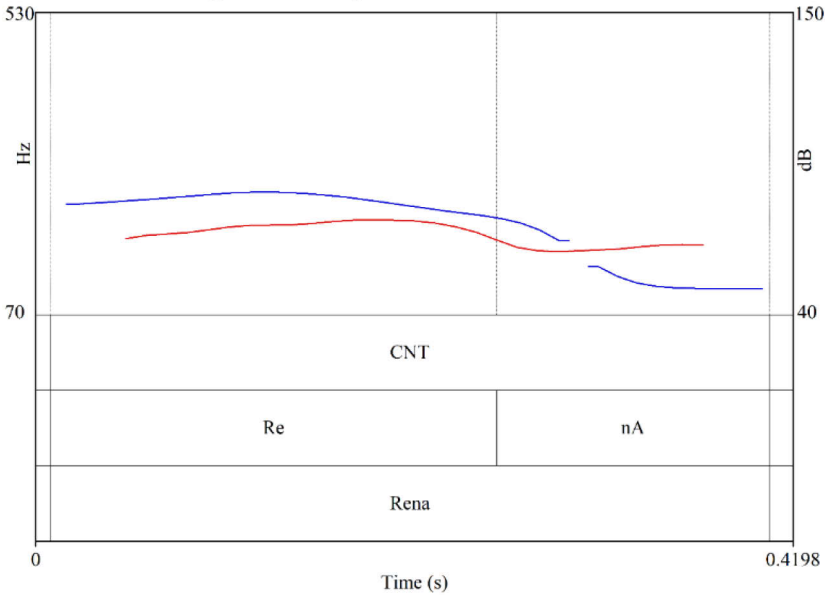
Figure 10 – Representation of the ALL unit of audio 17.



Source: Raso & Ferrari, 2020.

Example 18: [audio 18] bfamdl01[27]
FLA*: (<cê vai embora que> dia /) Rena //=CNT=

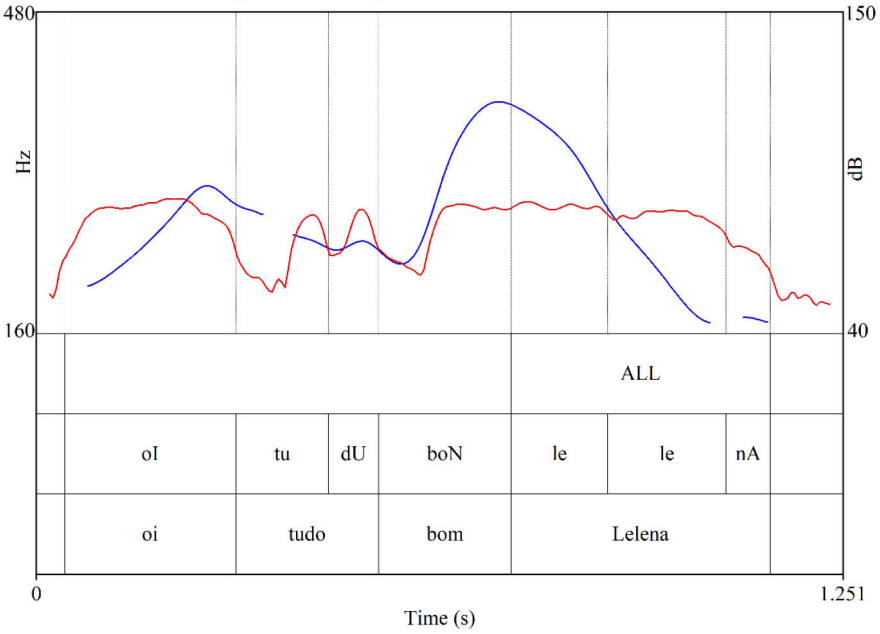
Figure 11 – Representation of audio 18.



Source: Raso & Ferrari, 2020.

Example 19: [audio 20]
*LUR: oi / tudo bom / Lelena //=ALL=
hello / everything OK / Lelena //

Figure 12 – Representation of audio 19.

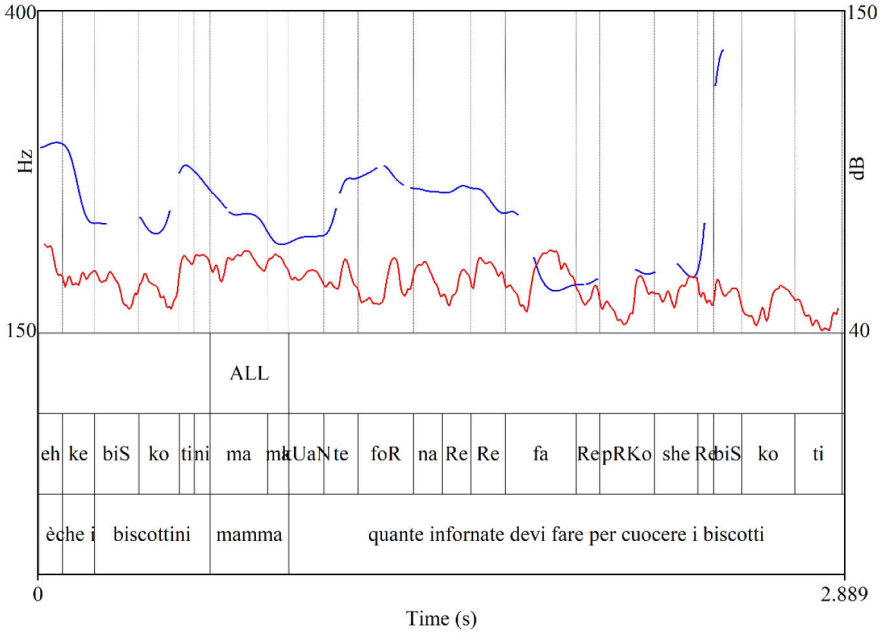


Fonte: Vieira & Raso, 2016.

Figure 12 can be compared with figure 6 since the same lexeme (Lelena), stressed on the second syllable, is performed as a CNT in example 13 (fig. 6) and as an ALL here.

Example 20: [audio 20] ifamdl14[134]
*VAL: è che i biscottini / mamma /=ALL= quante infornate devi fare per cuocere i biscotti //
the fact is that the cookies / mom / how many times should you put them in the oven to cook them //

Figure 13 – Representation of audio 20.



Source: Raso & Vieira, 2016.

Figure 13 shows the form of ALL when its position is not the final one, but it is inside the utterance. The unit has a falling movement that becomes flat and then falls at the end. This is even more evident in figure 4 (ex. 11), where the second unit (*Massimo*) is an ALL. This shows that ALL's form is influenced by its position in the utterance, while CNT is not.

4.2 Parameter extraction

To describe these changes in the shapes of DM contours, as described in the preceding section, we propose to compare two approaches. One, deriving prosodic descriptors from the raw measurements, has been tested by Gobbo (2019) for the classification of DMs. The other one, still untested on DMs, consists in extracting parameters describing the shapes of a curve fitting the raw F0 measurement; it has been tested, e.g., by Cavalcante (2020) for describing the topic unit.

4.2.1 Prosodic descriptors of prosody

The work presented in Gobbo (2019) is a good example of a prosodic feature set applied to the DM case. The author extracted from a corpus the following set of prosodic descriptors (details on their measurement in the original work):

- Five duration descriptors: the raw and normalized duration (in milliseconds) of the DM and of its syllables (mean of the raw and normalized duration), the number of syllables of the DM (normalized durations are obtained using BARBOSA, 2013, algorithm).
- Four intensity descriptors: the mean, standard deviation, minimum and maximum of intensity (in dB) level over the DM (calculated with reference to the comment unit's level).
- Eight F0 descriptors: the mean, standard deviation, minimum and maximum of F0 level (in semitones) over the DM, plus the F0 range, and F0 slopes (over the complete DM, until the final stress, from the final stress).
- Eight "alignment" parameters, indicating where stands the intensity or F0 maximum or minimum, in relation to the DM duration or the stress: relative position of the minimum or maximum of intensity or F0 within the DM

These 25 descriptors were then fed in multivariate data analysis procedures, which will be described in 4.3.

4.2.1 Parameters of shape description

On the basis of the raw F0 measurements, a curve fitting algorithm can be applied on the part of the DM that has to be taken into account. Within the FDA framework (RAMSAY; SILVERMAN, 2005), there is a need to find a single timeframe that fits all the units to be compared, so the timing of the items has to be comparable;

as the events of interests do not necessarily appear at the same position across items, a registration procedure allows the definition of anchor points between which comparable variations do occur. This process can, for instance, be used to match stress syllables across items of different metric structures (see CAVALCANTE, 2020). The curve fitting is then done minimizing a cost function (the roughness of the curve) that deals with the smoothness of the final curve – this process is similar to the MOMEL approach (HIRST; ESPESSER 1993) that fits a continuous spline across a set of anchor points and proved an adequate manner to obtain a close-copy stylization of the raw F0 points.

By doing so, we only use the intonation curve, and we consider F0 points as having an identical weight on the perception of speech melody. It may be possible to introduce weighting factors to put more importance on the passages of pitch that are produced with a stronger voice and thus shall bear a more important perceptual role. Building on Hermes (1998a, b), one may use the maximum amplitude of the subharmonic subspectrum as a weighting factor or another measure of loudness related to voicing (see, e.g., D'ALESSANDRO; RILLIARD; LE BEUX, 2011). Duration is also certainly an important factor to take into account so as to have a reliable prosodic representation. One problem with the duration being its measurement at a syllabic (or phonemic) level – but local speech rate estimations may be derived as a continuous curve (MIXDORFF; PFITZINGER, 2005): one may use it as a weighting factor or as a second dimension of the curve fitting (fitting then a surface; the interpretation may be more complex). The obtained models can be summarized by their parameters, which are used in the following steps.

4.3 Multivariate analyses

4.3.1 Classification

On the basis of his 25 prosodic descriptors, Gobbo (2019) applied multivariate methods so as to find the best combination of them able to separate the three labeled DMs categories (ALL, CNT, and INP). He found that a subset of three descriptors gave a global 74% accuracy for classifying these three categories. These descriptors were the mean intensity and F0, and the F0 slope up to the stressed syllable. Adding more descriptors led to higher accuracies (up to about 80%). However, by changing the set of selected parameters as we increase the number of parameters, one may get less reliable and interpretable results on unseen data.

Classification of DM's prosodic shapes has still not been done using FDA data (but is an ongoing process), but the feasibility of this approach has been demonstrated by Cavalcante (2020) for the forms of the Topic unit.

4.3.2 Qualitative analysis

On the same 25 prosodic descriptors (or potentially on the FDA parameters), one may pursue a descriptive analysis of the DMs that were not attributed a given function so as to regroup them on the basis of the prosodic characteristics to check if this leads to clusters that can then be interpreted at a functional level. Gobbo (2019) proposed a rapid overview of this possibility, sorting the AUX units[4] of his corpus into categories of shapes. A qualitative analysis of this data[5] reveals that, besides the forms

[4] Out of 564 tokens analyzed by Gobbo (2020) 414 tokens received the tag AUX so as to indicate that they were DMs (or Auxiliary Units) but that they could not, at that moment, be classified into a well-known category.

[5] Qualitative reanalysis of the 414 tokens left out of the main model proposed by Gobbo (2020). This work resulted in the regrouping of data in three main categories and in the exclusion of items due to hypersegmentation and quality issues (RASO; SANTOS, in preparation).

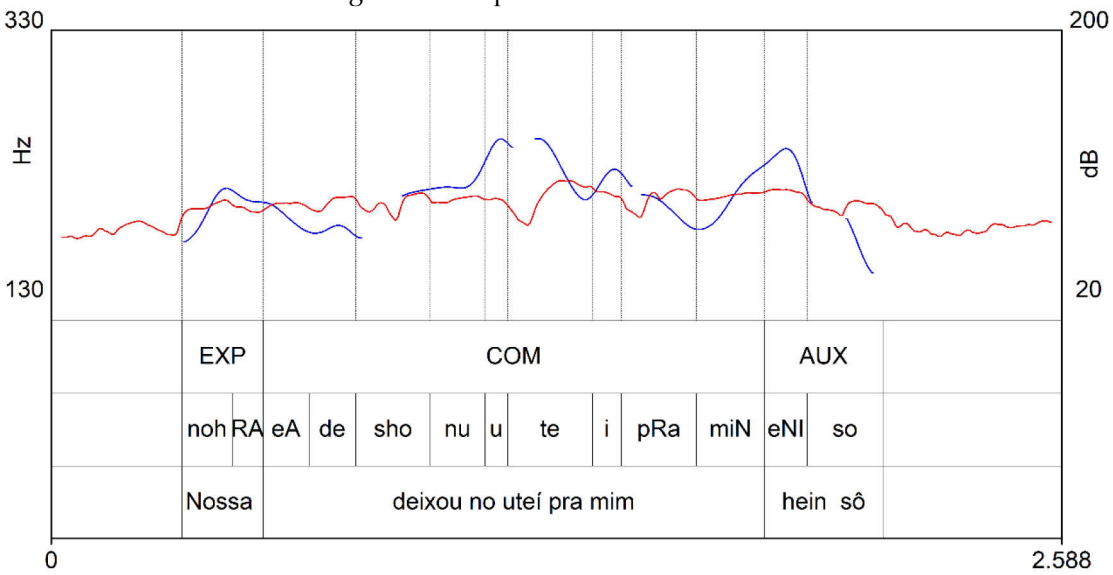
described for ALL, CNP, and INP, other three forms can be put forth that seem to bear prosodic and functional coherences. We give a brief overview of these forms, showing some prototypical examples and giving their functions. The reader is warned that this is an ongoing work and that a prosodic description (with acoustic measurements) is still needed in order to confirm the forms.

4.3.2.1 EXP

The first form is labeled under the tag EXP (Expressive), which is generally used to convey some surprise (but not as an illocution) or to give emotional support to the act being performed. It has a rising f0 shape on the stressed syllable, which can shortly fall in the presence of post-stress segmental material. There seems to be some lengthening on the stressed syllable (at least) with respect to the mean syllabic duration of COM (Comment unit), although this aspect needs to be confirmed. Its intensity seems to be on the same level as that of COM or a bit lower. It is frequently found at the very beginning of the terminated sequence (utterance/stanza) or at the beginning of reported speech.

Example 21: [audio 21] bfamcv03[138]
*TON: Nossa /=EXP= ea deixou no uteí pra mim /=COM= hein sô //=CNT=
Holy /=EXP= she left it in the UCI for me / you saw //

Figure 14 – Representation of audio 21.



Source: the authors.

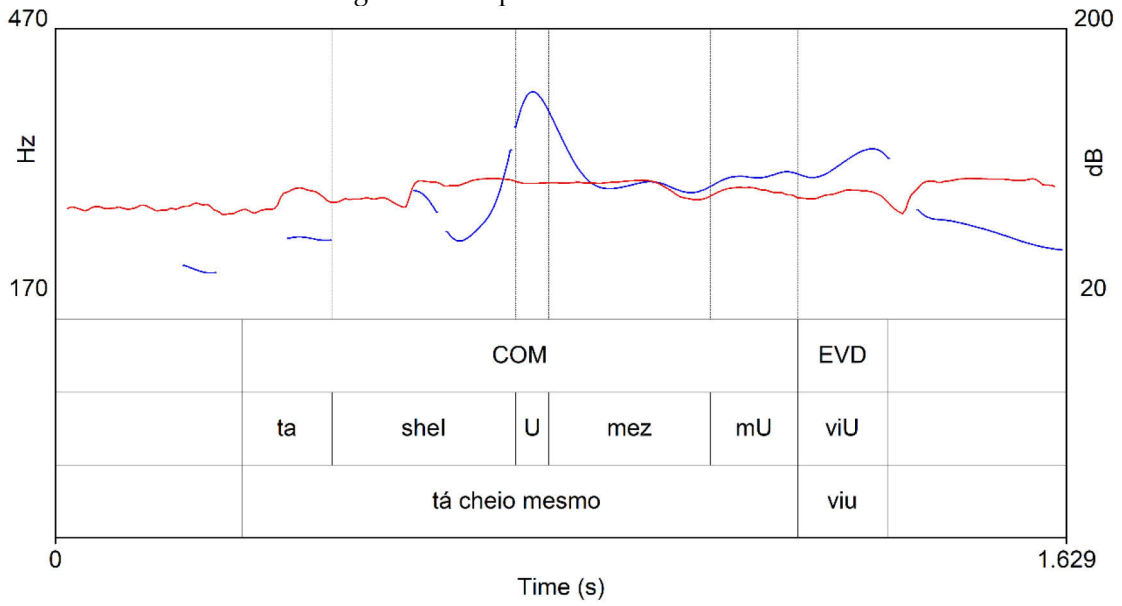
Frequent lexical fillers are *Nossa/No'* (Holy), *ah* (oh), and *não* (no).

4.3.2.2 EVD

The second form is tentatively tagged as EVD (Evidentiator - not to be confounded with grammatical markers of evidentiality). It can highlight what was said and secure, at the same time, the addressee's attention. It is characterized by a (generally) slightly rising f0 shape, lower intensity with respect to COM, and syllabic duration that tends to be shorter than that of COM. The slope of the rising movement can vary from almost flat to markedly slopy. All the same, it suffices to create a contrast with a falling f0 movement with the same f0 mean. This effect seems to be in connection with the very low intensity it is often produced with.

Example 22: [audio 22] bfamdl01[199]
*REN: tá cheio mesmo /=COM= viu //=EVD=
it's really crowded /=COM= huh //=EVD=

Figure 15 – Representation of audio 22.



Source: the authors.

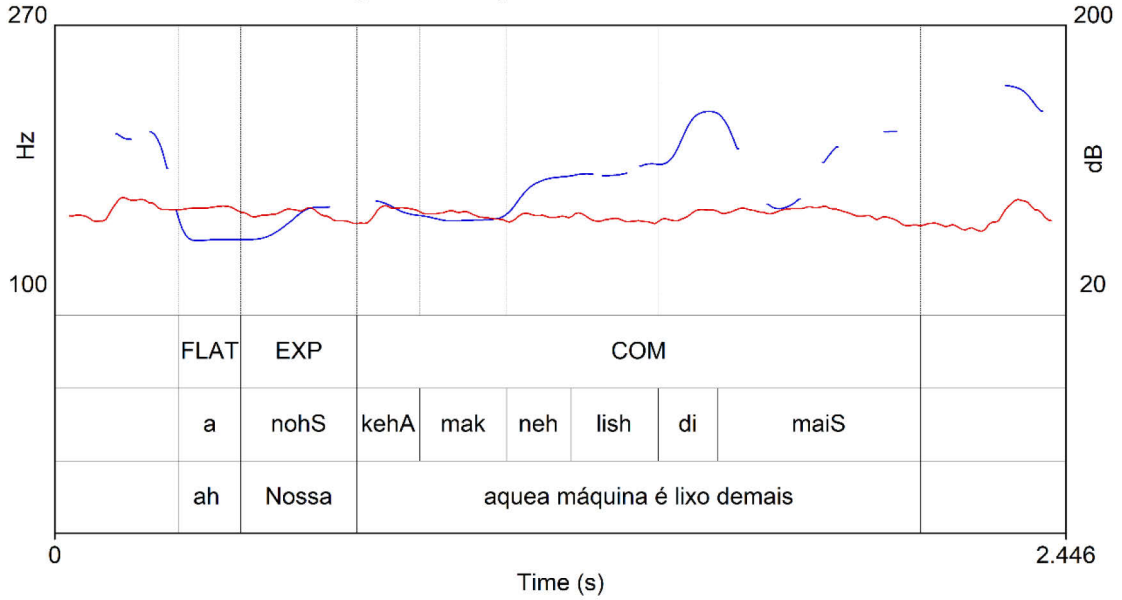
Frequent lexical items fulfilling this form are *né* (isn't it), *hein* (huh), *viu* (you see), and *sabe* (you know).

4.3.2.3 Initial flat form

The last form seems to convey a coherent function, but more data are needed for a better functional definition. We hypothesize that it may be used as a sign that the speaker has just realized something about what s/he is about to say. As to its prosodic form, this group has, as indicated by the tentative tag, a flat f0 profile, syllabic mean duration shorter than that of COM, and generally mean to high intensity levels. We found it exclusively in the initial position.

Example 23: [audio 23] bfammn05[102]
*JUN: ah /=FLAT= Nossa /=EXP= aquea máquina é lixo demais //=COM=
oh /=FLAT= Holy / that camera is complete trash //

Figure 16 – Representation of audio 23.



Source: the authors.

Frequent lexical fillers found for this form are *ah* (oh), *é* (yes), and *não* (no). It is noteworthy that the degree of pragmaticalization of these items can be shown by the fact that *yes* and *no* can be interchanged without detriment to the meaning of the utterance.

5 Conclusions & perspectives

In this paper, we propose a new way to identify DMs and their specific functions based on prosodic parameters after discussing different aspects of the mainstream literature on this topic. Prosody allows accounting for the main formal features that convey both the function of being a DM and the specific functions performed by different kinds of DMs. We showed how and why it is possible to establish when a lexical item behaves as DM, consequently giving a clear functional and formal definition of what should be considered a DM. We also described the specific formal prosodic cues that allow us to recognize the specific function of the

different types of DMs: at this point, we think that there are probably six types of DMs with interactional functions, for five of which we presented a more reliable picture.

We presented the methodology applied to the description of the different prosodic cues and the steps we have followed so far and will follow in order to automatically extract from new data the different kinds of DMs, discussing different strategies and the pros and contras of each of them.

The contribution of the paper is, therefore, twofold: it proposes a new linguistic solution to the problem of DM identification, and it shows how to model them in order to turn possible an automatic extraction of each specific functional item.

References

AIJMER, K. **Understanding pragmatic markers**: a variational pragmatic approach. Edinburgh: Edinburgh Univ. Press, 2013. DOI https://doi.org/10.1515/9780748635511

AIJMER, K.; SIMON-VANDENBERGEN, A.-M. (ed.). **Pragmatic markers in contrast**. Amsterdam: Elsevier, 2006. DOI https://doi.org/10.1163/9780080480299

D'ALESSANDRO, C. Voice source parameters and prosodic analysis. *In*: SUDHOFF, S.; LENERTOVA, D.; MEYER, R.; PAPPERT, S.; AUGURZKY, P.; MLEINEK, I.; RICHTER, N.; SCHLIESSER, J. (ed.). **Methods in empirical prosody research**. Berlin: Walter de Gruyter, 2006. p. 63–87.

D'ALESSANDRO, C.; MERTENS, P. Automatic pitch contour stylization using a model of tonal perception. **Computer Speech & Language**, vol. 9, no. 3, p. 257–288, 1995. DOI https://doi.org/10.1006/csla.1995.0013

D'ALESSANDRO, C.; RILLIARD, A.; LE BEUX, S. Chironomic stylization of intonation. **Journal of the Acoustical Society of America**, vol. 129, no. 3, p. 1594–1604, 2011. DOI https://doi.org/10.1121/1.3531802

D'ALESSANDRO, C.; FEUGÈRE, L.; LE BEUX, S.; PERROTIN, O.; RILLIARD, A. Drawing melodies: Evaluation of chironomic singing synthesis. **The Journal of the Acoustical Society of America**, vol. 135, no. 6, p. 3601–3612, 2014. DOI https://doi.org/10.1121/1.4875718

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

BALLY, C. **Linguistique générale et linguistique française**. 3rd ed. Berne: A. Francke, 1950.

BANSE, R.; SCHERER, K. R. Acoustic profiles in vocal emotion expression. **Journal of Personality and Social Psychology**, vol. 70, p. 614–636, 1996. DOI https://doi.org/10.1037/0022-3514.70.3.614

BARBOSA, P. A. From syntax to acoustic duration: A dynamical model of speech rhythm production. **Speech Communication**, vol. 49, no. 9, p. 725–742, 2007. DOI https://doi.org/10.1016/j.specom.2007.04.013

BARBOSA, P. A. Semi-automatic and automatic tools for generating prosodic descriptors for prosody research. **Proceedings from TRASP**, p. 86–90, 2013.

BAZZANELLA, C. Phatic connectives as interactional cues in contemporary spoken Italian. **Journal of Pragmatics**, vol. 14, no. 4, p. 629–647, Aug. 1990. DOI https://doi.org/10.1016/0378-2166(90)90034-B

BOERSMA, P.; WEENINK, D. **Praat**: doing phonetics by computer [Computer program]. Version 6.1.16, 2020. Available at: http://www.praat.org/.

BOLDEN, G. B. Discourse Markers. *In*: TRACY, K.; SANDEL, T.; ILIE, C. (ed.). **The International Encyclopedia of Language and Social Interaction**. 1st ed. Wiley, 2015. p. 1–7.

BRINTON, L. J. **Pragmatic Markers in English**: Grammaticalization and Discourse Functions. De Gruyter Mouton, 1996. DOI https://doi.org/10.1515/9783110907582

CAMACHO, A.; HARRIS, J. G. A sawtooth waveform inspired pitch estimator for speech and music. **The Journal of the Acoustical Society of America**, vol. 124, no. 3, p. 1638–1652, Sep. 2008. DOI https://doi.org/10.1121/1.2951592

CAMPBELL, W.N.; ISARD, S.D. Segment durations in a syllable frame. **Journal of Phonetics**, vol. 19, no. 1, p. 37–47, Jan. 1991. DOI https://doi.org/10.1016/S0095-4470(19)30315-8

CAVALCANTE, F. A. **The information unit of topic**: a crosslinguistic, statistical study based on spontaneous speech corpora. 2020. PhD Thesis – Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, 2020.

COLE, J. Prosody in context: a review. **Language, Cognition and Neuroscience**, vol. 30, no. 1–2, p. 1–31, 7 Feb. 2015. DOI https://doi.org/10.1080/23273798.2014.963130

CONTINI, M.; PROFILI, O. L'intonation de l'italien régional - un modèle de description par traits. 1989. **Mélanges de phonétique générale et expérimentale offerts à Péla Simon**. Strasbourg: Publications de l'Institut de phonétique de Strasbourg, 1989. p. 855–870.

CRESTI, E. **Corpus di italiano parlato**. Firenze: presso l'Accademia della Crusca, 2000.

CRESTI, E. La stanza: Un'unità di costruzione testuale del parlato. *In*: SILFI 2008, 2010. **Atti del X Congresso della Società Internazionale di Linguistica e Filologia Italiana**. Basel: Firenze: Cesati, 2010. p. 713–732.

CRESTI, E.; MONEGLIA, M. (ed.). **C-ORAL-ROM**: Integrated Reference Corpora for Spoken Romance Languages. vol. 15 (Studies in Corpus Linguistics). Amsterdam: John Benjamins Publishing Company, 2005. DOI https://doi.org/10.1075/scl.15

CRESTI, E.; MONEGLIA, M. The Discourse Connector according to the Language into Act Theory: data from IPIC Italiano. *In*: BIDESE, E.; CASALICCHIO, J.; MORONI, M. C. (ed.). **La linguistica vista dalle Alpi Linguistic views from the Alps**. Berlin: Peter Lang Verlag, 2020.

DAVIDSON, L. The Effects of Pitch, Gender, and Prosodic Context on the Identification of Creaky Voice. **Phonetica**, vol. 76, no. 4, p. 235–262, 1 Jul. 2019. DOI https://doi.org/10.1159/000490948

DE CASTRO MOUTINHO, L.; COIMBRA, R. L.; RILLIARD, A.; ROMANO, A. Mesure de la variation prosodique diatopique en portugais européen. **Estudios de fonética experimental**, vol. 20, p. 33–55, 2011.

DE CHEVEIGNÉ, A.; KAWAHARA, H. YIN, a fundamental frequency estimator for speech and music. **The Journal of the Acoustical Society of America**, vol. 111, no. 4, p. 1917–1930, 2002. DOI https://doi.org/10.1121/1.1458024

DEGAND, L. 7 'So very fast then' Discourse Markers at Left and Right Periphery in Spoken French. *In*: BEECHING, K.; DETGES, U. (ed.). **Discourse Functions at the Left and Right Periphery**, BRILL, 2014. p. 151–178. DOI https://doi.org/10.1163/9789004274822_008

DI CRISTO, A.; HIRST, DJ Modelling French Micromelody: Analysis and Synthesis. **Phonetica**, vol. 43, no. 1–3, p. 11–30, 1 Jan. 1986. DOI https://doi.org/10.1159/000261758

DUBĚDA, T.; KELLER, E. Microprosodic aspects of vowel dynamics—an acoustic study of French, English and Czech. **Journal of Phonetics**, vol. 33, no. 4, p. 447–464, Oct. 2005. DOI https://doi.org/10.1016/j.wocn.2005.02.003

EVANS, J.; CHU, M.; ASTON, J. A. D.; SU, C. Linguistic and human effects on F0 in a tonal dialect of Qiang. **Phonetica**, vol. 67, no. 1–2, p. 82–99, 2010. DOI https://doi.org/10.1159/000319380

EYBEN, F; SCHULLER, B. openSMILE:): the Munich open-source large-scale multimedia feature extractor. **ACM SIGMultimedia Records**, vol. 6, no. 4, p. 4–13, 28 Jan. 2015. DOI https://doi.org/10.1145/2729095.2729097

EYBEN, F.; SCHERER, K. R.; SCHULLER, B. W.; SUNDBERG, J.; ANDRE, E.; BUSSO, C.; DEVILLERS, L. Y.; EPPS, J.; LAUKKA, P.; NARAYANAN, S. S.; TRUONG, K. P. The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing. **IEEE Transactions on Affective Computing**, vol. 7, no. 2, p. 190–202, 1 Apr. 2016. DOI https://doi.org/10.1109/TAFFC.2015.2457417

FISCHER, K. (ed.). **Approaches to discourse particles**. Studies in pragmatics, 1. Oxford: Elsevier, 2006a. DOI https://doi.org/10.1163/9780080461588

FISCHER, K. Towards an understanding of the spectrum of approaches to discourse particles: introduction to the volume. *In*: FISCHER, K. (ed.). **Approaches to discourse particles**. Studies in pragmatics. Oxford: Elsevier, 2006b. p. 1–20. DOI https://doi.org/10.1163/9780080461588_002

FRANK-JOB, B. A dynamic-interactional approach to discourse markers. In: FISCHER, Kerstin (ed.). **Approaches to discourse particles**. Studies in pragmatics. Oxford: Elsevier, 2006. p. 395–413.

FROSALI, F. Il lessico degli Ausili Dialogici. *In*: CRESTI, E. (ed.). Prospettive nello studio del lessico italiano: **Atti del IX Congresso SILFI**. Proceedings e report. 1st ed. Florence: Firenze University Press, 2008. vol. 40, p. 417–424.

FUJISAKI, H. Dynamic characteristics of voice fundamental frequency in speech and singing. In: MACNEILAGE, P. (ed.). **The production of speech**. New York, NY: Springer, 1983. p. 39–55. DOI https://doi.org/10.1007/978-1-4613-8202-7_3

FUJISAKI, H. A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. *In*: FUJIMURA, O. (ed.). **Vocal fold physiology**: voice production, mechanisms and functions. New York, NY: Raven, 1988. p. 347–355.

GERRATT, B. R.; KREIMAN, J. Toward a taxonomy of non-modal phonation. **Journal of Phonetics**, vol. 29, no. 4, p. 365–381, Oct. 2001. DOI https://doi.org/10.1006/jpho.2001.0149

GOBBO, O. **Marcadores discursivos em uma perspectiva informacional**: análise prosódica e estatística. 2019. Master Thesis – Federal University of Minas Gerais, Belo Horizonte, Brazil, 2019.

GOUDBEEK, M.; SCHERER, K. Beyond arousal: Valence and potency control cues in the vocal expression of emotion. **The Journal of the Acoustical Society of America**, vol. 128, no. 3, p. 1322–1336, 2010. DOI https://doi.org/10.1121/1.3466853

GRABE, E.; KOCHANSKI, G.; COLEMAN, J. Connecting intonation labels to mathematical descriptions of fundamental frequency. **Language and speech**, vol. 50, no. 3, p. 281–310, 2007. DOI https://doi.org/10.1177/00238309070500030101

GURLEKIAN, J.; MIXDORFF, H.; EVIN, D.; TORRES, H.; PFITZINGER, H. R. Alignment of F0 model parameters with final and non-final accents in Argentinean Spanish. 2010. **Proceedings**. Speech Prosody 2010. Chicago, USA: 2010. p. paper 131.

HADJIPANTELIS, P. Z; ASTON, J. A. D.; EVANS, J. P. Characterizing fundamental frequency in Mandarin: A functional principal component approach utilizing mixed effect models. **The Journal of the Acoustical Society of America**, vol. 131, no. 6, p. 4651–4664, 2012. DOI https://doi.org/10.1121/1.4714345

HAMMARBERG, B.; FRITZELL, B.; GAUFIN, J.; SUNDBERG, J.; WEDIN, L. Perceptual and Acoustic Correlates of Abnormal Voice Qualities. **Acta Oto-**

Laryngologica, vol. 90, no. 1–6, p. 441–451, Jan. 1980. DOI https://doi.org/10.3109/00016488009131746

HERMES, D. J. Auditory and Visual Similarity of Pitch Contours. **Journal of Speech**, Language, and Hearing Research, vol. 41, no. 1, p. 63–72, 1998a. DOI https://doi.org/10.1044/jslhr.4101.63

HERMES, D. J. Measuring the Perceptual Similarity of Pitch Contours. **Journal of Speech, Language, and Hearing Research**, vol. 41, no. 1, p. 73–82, 1998b. DOI https://doi.org/10.1044/jslhr.4101.73

HIRST, D.; DI CRISTO, A. A survey of intonation systems. *In*: HIRST, D.; DI CRISTO, A. (ed.). **Intonation systems**: a survey of twenty languages. Cambridge, U.K.: Cambridge University Press, 1998. p. 1–44.

HIRST, D; DI CRISTO, A.; ESPESSER, R. Levels of Representation and Levels of Analysis for the Description of Intonation Systems. *In*: HORNE, M. (ed.). **Prosody**: Theory and Experiment. Text, Speech and Language Technology. vol. 14. Dordrecht: Springer Netherlands, 2000. p. 51–87. DOI https://doi.org/10.1007/978-94-015-9413-4_4

HIRST, D.; ESPESSER, R. Automatic Modelling of Fundamental Frequency Using a Quadratic Spline Function. **Travaux de l'Institut de Phonétique d'Aix**, vol. 15, p. 75–85, 1993.

HIRST, D.; RILLIARD, A.; AUBERGÉ, V. Comparison of subjective evaluation and an objective evaluation metric for prosody in text-to-speech synthesis. **The Third ESCA/COCOSDA Workshop (ETRW) on Speech Synthesis**. 1998.

HONDA, K. Physiological factors causing tonal characteristics of speech: from global to local prosody. 2004. **International Conference on Speech Prosody 2004**. Nara, Japan: 2004. p. 739–744.

KOCHANSKI, G.; SHIH, C. Prosody modeling with soft templates. **Speech Communication**, vol. 39, no. 3–4, p. 311–352, Feb. 2003. DOI https://doi.org/10.1016/S0167-6393(02)00047-X

KOHLER, K. J. What is emphasis and how is it coded? 2006. **Proceedings**. Speech Prosody 2006. Dresden, Germany: 2006. paper 015.

KREIMAN, J.; GERRATT, B. R.; ANTOÑANZAS-BARROSO, N. Measures of the Glottal Source Spectrum. **Journal of Speech, Language, and Hearing Research**, vol. 50, no. 3, p. 595–610, Jun. 2007. https://doi.org/10.1044/1092-4388(2007/042)

LAI, C. Interpreting final rises: task and role factors. **7th International Conference on Speech Prosody**, Dublin, Ireland, 2014. p. 520–524.

LEVITT, H.; RABINER, L. R. Analysis of fundamental frequency contours in speech. **The Journal of the Acoustical Society of America**, vol. 49, p. 569, 1971. DOI https://doi.org/10.1121/1.1912388

LIÉNARD, J.-S.; DI BENEDETTO, M.-G. Effect of vocal effort on spectral properties of vowels. **The Journal of the Acoustical Society of America**, vol. 106, no. 1, p. 411–422, Jul. 1999. DOI https://doi.org/10.1121/1.428140

LIÉNARD, J.-S. Quantifying vocal effort from the shape of the one-third octave long-term-average spectrum of speech. **The Journal of the Acoustical Society of America**, vol. 146, no. 4, p. EL369–EL375, Oct. 2019. DOI https://doi.org/10.1121/1.5129677

LÖFQVIST, A.; MANDERSSON, B. Long-Time Average Spectrum of Speech and Voice Analysis. **Folia Phoniatrica et Logopaedica**, vol. 39, no. 5, p. 221–229, 1987. DOI https://doi.org/10.1159/000265863

MACARY, M.; TAHON, M.; ESTEVE, Y.; ROUSSEAU, A. On the Use of Self-Supervised Pre-Trained Acoustic and Linguistic Features for Continuous Speech Emotion Recognition. *In*: **IEEE SPOKEN LANGUAGE TECHNOLOGY WORKSHOP (SLT)**, Shenzhen, China, 2021. p. 373–380. DOI https://doi.org/10.1109/SLT48900.2021.9383456

MAIA ROCHA, B.; RASO, T. A unidade informacional de Introdutor Locutivo no português do Brasil: uma primeira descrição baseada em corpus. **Domínios de Lingu@gem**, vol. 5, no. 1, p. 327–343, Jul. 2011. Available at: https://seer.ufu.br/index.php/dominiosdelinguagem/article/view/12479

MARTIN, Ph. Les problèmes de l'intonation: recherches et applications. **Langue française**, vol. 19, no. 1, p. 4–32, 1973. DOI https://doi.org/10.3406/lfr.1973.5638

MARTIN, Ph. Prosodic and rhythmic structures in French. **Linguistics**, vol. 25, no. 5, p. 925–950, 1987.

MARTIN, Ph. Multi methods pitch tracking. *In*: **Speech Prosody**, 2012. Shanghai, China, 2012, p. 47–50.

MERTENS, P. The prosogram: semi-automatic transcription of prosody based on a tonal perception model. **International Conference on Speech Prosody 2004**, Nara, Japan, 2004, p. 549–552. DOI https://doi.org/10.20396/joss.v4i2.15053

MERTENS, P. Polytonia: a system for the automatic transcription of tonal aspects in speech corpora. **Journal of Speech Sciences**, vol. 4, no. 2, p. 17–57, 5 Feb. 2014.

MEUNIER, S.; CHATRON, J.; ABS, B.; PONSOT, E.; SUSINI, P. Effect of Pitch on the Asymmetry in Global Loudness Between Rising- and Falling-Intensity Sounds. **Acta Acustica united with Acustica**, vol. 104, no. 5, p. 770–773, 1 Sep. 2018. DOI https://doi.org/10.3813/AAA.919220

MITTMANN, M. M. **O C-ORAL-BRASIL e o estudo da fala informal**: um novo olhar sobre o tópico no português brasileiro. 248 f. PhD Thesis – Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, 2012.

MIXDORFF, H. A novel approach to the fully automatic extraction of Fujisaki model parameters. Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. **Proceedings**. 2000 IEEE International Conference on Speech and Signal Processing, IEEE, 2000. vol. 3. p. 1281–1284.

MIXDORFF, H.; PFITZINGER, H. R. Analysing fundamental frequency contours and local speech rate in map task dialogs. **Speech Communication**, vol. 46, no. 3–4, p. 310–325, Jul. 2005. DOI https://doi.org/10.1016/j.specom.2005.02.019

MONEGLIA, M.; RASO, T. Appendix: Notes on the Language into Act Theory. *In*: RASO, T.; MELLO, H. (eds.). **Studies in Corpus Linguistics**. Amsterdam: John Benjamins Publishing Company, 2014. vol. 61, p. 468–495. DOI https://doi.org/10.1075/scl.61.15mon

MOORE, B. C. J.; GLASBERG, B. R.; VARATHANATHAN, A.; SCHLITTENLACHER, J. A Loudness Model for Time-Varying Sounds Incorporating Binaural Inhibition. **Trends in Hearing**, vol. 20, Jan. 2016. DOI https://doi.org/10.1177/2331216516682698

NERBONNE, J.; COLEN, R.; GOOSKENS, C. S.; LEINONEN, T.; KLEIWEG, P. Gabmap – A web application for dialectology. **Dialectologia**, vol. SI II, p. 65–89, 2011.

NORDENBERG, M.; SUNDBERG, J. Effect on LTAS of vocal loudness variation. **Logopedics Phoniatrics Vocology**, vol. 29, no. 4, p. 183–191, Dec. 2004. DOI https://doi.org/10.1080/14015430410004689

PIERREHUMBERT, J. B. **The phonology and phonetics of English intonation**. PhD Thesis. Massachusetts Institute of Technology, 1980.

PIERREHUMBERT, J. Synthesizing intonation. **The Journal of the Acoustical Society of America**, vol. 70, no. 4, p. 985–995, Oct. 1981. DOI https://doi.org/10.1121/1.387033

RAMSAY, J. O.; SILVERMAN, B. W. **Functional data analysis**. 2nd ed. New York: Springer, 2005. DOI https://doi.org/10.1007/b98888

RASO, T. Prosodic constraints for discourse markers. *In*: RASO, T.; MELLO, H. (ed.). **Studies in Corpus Linguistics**. vol. 61. Amsterdam: John Benjamins Publishing Company, 2014. p. 411–467. DOI https://doi.org/10.1075/scl.61.14ras

RASO, T.; CAVALCANTE, F. A.; MITTMANN, M. M. Prosodic forms of the Topic information unit in a cross–linguistic perspective. A first survey. *In*: DE MEO, A.; DOVETTO, F. M. (ed.). **La comunicazione parlata / Spoken Communication**. Rome: Aracne, 2017. p. 473–498.

RASO, T.; FERRARI, L. A. Uso dei Segnali Discorsivi in corpora di parlato spontaneo italiano e brasiliano. *In*: FERRONI, R.; BIRELLO, M. (ed.). **La competenza discorsiva e interazionale: a lezione di lingua straniera**. Canterano (Roma): Aracne, 2020. p. 61–107.

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

RASO, T.; VIEIRA, M. A. A description of Dialogic Units/Discourse Markers in spontaneous speech corpora based on phonetic parameters. **CHIMERA**: Romance Corpora and Linguistic Studies, vol. 3, no. 2, p. 221–249, 2016.

RENCHER, A. C.; CHRISTENSEN, W. F. **Methods of multivariate analysis**. Third Edition. Hoboken, New Jersey: Wiley, 2012. DOI https://doi.org/10.1002/9781118391686

RILLIARD, A. Geoprosody – Quantitative approaches of prosodic variation across dialects. *In*: VIEIRA, M. S. M.; WIEDEMER, M. L. (ed.). **Dimensões e Experiências em**

Sociolinguística. São Paulo: Editora Blucher, 2019. p. 55–83. DOI https://doi.org/10.5151/9788521218746-02

ROCHA, B.; MELLO, H.; RASO, T. Para a compilação do C-ORAL-ANGOLA. **Filologia e Linguística Portuguesa**, vol. 20, no. Especial, p. 139–157, 30 Dec. 2018. DOI https://doi.org/10.11606/issn.2176-9419.v20iEspecialp139-157

ROMERO-TRILLO, J. Discourse Markers. *In*: MEY, J. (ed.). **Concise encyclopedia of pragmatics**. Amsterdam; New York: Elsevier, 1998. p. 191–194.

ROSSI, M. Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole. **Phonetica**, vol. 23, no. 1, p. 1–33, 1971. DOI https://doi.org/10.1159/000259328

SCHIFFRIN, D. **Discourse Markers**. Cambridge University Press, 1987. DOI https://doi.org/10.1017/CBO9780511611841

SCHOURUP, L. Discourse markers. **Lingua**, vol. 107, no. 3–4, p. 227–265, Apr. 1999. DOI https://doi.org/10.1016/S0024-3841(96)90026-1

SIGNOL, F.; BARRAS, C.; LIENARD, J.-S. Evaluation of the Pitch Estimation Algorithms in the monopitch and multipitch cases. *In*: **Proceedings of Acoustics'08**. Paris, France: May 2008. p. 675–680.

SIGNORINI, S. **Topic e soggetto in corpora di italiano parlato spontaneo**. 2005. PhD Thesis. Università di Firenze, Firenze, Italy, 2005.

SUNDBERG, J.; NORDENBERG, M. Effects of vocal loudness variation on spectrum balance as reflected by the alpha measure of long-term-average spectra of speech. **The Journal of the Acoustical Society of America**, vol. 120, no. 1, p. 453–457, Jul. 2006. DOI https://doi.org/10.1121/1.2208451

ŠVEC, J. G.; GRANQVIST, S. Tutorial and Guidelines on Measurement of Sound Pressure Level in Voice and Speech. **Journal of Speech, Language, and Hearing Research**, vol. 61, no. 3, p. 441–461, 15 Mar. 2018. DOI https://doi.org/10.1044/2017_JSLHR-S-17-0095

'T HART, J.; COLLIER, R.; COHEN, A. **A perceptual study of intonation**: an experimental-phonetic approach to speech melody. Cambridge: Cambridge University Press, 1990. DOI https://doi.org/10.1017/CBO9780511627743

TALKIN, D. A robust algorithm for pitch tracking (RAPT). *In*: KLEIJN, W. Bastiaan; PALIWAL, K. K. (ed.). **Speech coding and synthesis**. Amsterdam: Elsevier, 1995. p. 495–518.

TITZE, I. R.; SUNDBERG, J. Vocal intensity in speakers and singers. **The Journal of the Acoustical Society of America**, vol. 91, no. 5, p. 2936–2946, May 1992. DOI https://doi.org/10.1121/1.402929

TRAUGOTT, E. C. Intersubjectification and clause periphery. **English Text Construction**, vol. 5, no. 1, p. 7–28, 4 May 2012. DOI https://doi.org/10.1075/etc.5.1.02trau

TRAUNMÜLLER, H.; ERIKSSON, A. Acoustic effects of variation in vocal effort by men, women, and children. **The Journal of the Acoustical Society of America**, vol. 107, no. 6, p. 3438–3451, Jun. 2000. DOI https://doi.org/10.1121/1.429414

TUCCI, I. L'inciso: caratteristiche morfosintattiche e intonative in un corpus di riferimento. *In*: ALBANO LEONI, F.; SENZA PELUSO, M. (ed.). **Il parlato italiano**: atti del convegno nazionale (Napoli, 13-15 febbraio 2003). Napoli: M. D'Auria editore, 2004. p. 1–14.

TUCCI, I. «Obiter dictum» La funzione informativa delle unità parentetiche. *In*: PETTORINO, M.; GIANNINI, A.; DOVETTO, F. M. (eds.). **La comunicazione parlata 3**. (Napoli, 23-25 febbraio 2009). Napoli: Liguori, 2010. p. 635–654.

VENABLES, W N; RIPLEY, B D. **Modern Applied Statistics with S**. Fourth edition. Springer, 2002. DOI https://doi.org/10.1007/978-0-387-21706-2

VINCENT, D.; ROSEC, O.; CHONAVEL, T. Glottal Closure Instant Estimation using an Appropriateness Measure of the Source and Continuity Constraints. *In*: **2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings**. vol. 1. Toulouse, France: IEEE, 2006. p. I-381-I–384.

XU, Y. Speech melody as articulatorily implemented communicative functions. **Speech Communication**, vol. 46, no. 3–4, p. 220–251, Jul. 2005. DOI https://doi.org/10.1016/j.specom.2005.02.014

Article received in: 12.21.2021 Article approved in: 21.03.2022