A description of Dialogic Units/Discourse Markers in spontaneous speech corpora based on phonetic parameters

Tommaso Raso, Marcelo A. Vieira
Universidade Federal de Minas Gerais

In this paper we propose a different way to face the notion of Discourse Markers (DM), based on prosodic features and using data from spontaneous speech corpora. We show how it is possible to predict DMs through prosodic parameters if we use corpora segmented into tone and information units. We will also show that DMs with different functions feature different prosodic profiles that are the formal counterpart of function in speech. Besides this, we will present the result of a first attempt of a statistical test, and discuss the limits and perspectives of the proposal. Finally, we will present a detailed description of four different DMs.

Keywords: Discourse Markers, Information Structure, Prosody, Corpora

1. Taking prosody into account may solve theoretical problems in Discourse Markers identification and functional analysis

The notion of Discourse Marker (DM) varies across frameworks (Schourup 1999; Fischer 2006; Traugott 2007; Bazzanella *et al.* 2008; Bolden 2015), but there is usually agreement about some characteristics that a lexical item should have in order to be considered a DM:

- a DM does not make up the text of the utterance, i.e. it does not have syntactic and semantic compositionality with rest of the utterance;
- the lexical item that functions as DM loses (at least part of) its semantic value and acquires some pragmatic value;
- DMs usually have free distribution.

CHIMERA. Romance Corpora and Linguistic Studies 1 (2016), 221-249. ISSN 2386-2629

Nevertheless, two important theoretical problems are always left unsolved:

- how to predict a DM, distinguishing it from a lexical item that does not function as a DM;
- which formal features can allow to identify the specific function of a given DM.

In our opinion, despite some partial attempts to look for different formal features, the two main reasons that cause the impossibility to answer these two questions are:

a. a general misunderstanding of how speech is structured, which leads to a lack of observation of the prosodic parameters and the way they work in order to convey functions;
b. an overestimation of lexicon, which leads to approach DMs as if their lexical form were the main feature to understand their function(s).

The result is that scholars cannot find formal features that can predict out of a specific context when a lexical item (or a small lexical sequence) functions as a DM and what its functions is. They need to analyze the context and decide, not without doubts and different possible interpretations, whether an item is or is not a DM, and, when they think it is, which possible function can be conveyed by it in that specific context. In the literature we also find many different proposals of DM functions, with models that reach even more than 20 different functions, involving many linguistic categories, including illocution, modality and attitude[1].

We will show that, looking at prosodic features, we can better account for both the above-mentioned problems: how to predict when a lexical item is a DM and how to distinguish among different functions.

In order to support our view, we will argue in 2 about the importance of prosodic segmentation of speech and how to use prosody for this goal; in 3 we will use prosodic features to show how DMs can be predicted; in 4 we will describe prosodic forms that can account for different functions of DMs. In 5 we will discuss the limits of our work so far and present the next steps necessary for its verification and completion.

[1] For more about these categories, besides what is said in the next sections, see Mello & Raso (2011) for our position.

Our data are taken from two balanced and comparable spontaneous speech minicorpora (Mittmann & Raso 2011; Panunzi & Mittmann 2014) extracted from C-ORAL-BRASIL (Raso & Mello 2012) and Italian C-ORAL-ROM (Cresti & Moneglia 2005) corpora.

2. Speech segmentation and unit of reference

One important theoretical question is how to segment speech into units of reference higher than the word level. This is a decisive point in order to understand and analyze linguistic relations. Here, we follow the proposal of the Language into Act Theory (Cresti 2000; Moneglia 2005; Moneglia & Raso 2014). For example, if we have a sequence in Brazilian Portuguese (BP) like

(1) João vai pro Rio até amanhã
‘João go(es) to Rio no later than tomorrow (*or* see you tomorrow)’

we can imagine different segmentations with different meanings, such as the following ones, among others[2]:

(2) João vai pro Rio // até amanhã //
‘João will go to Rio // see you tomorrow (*or* no later than tomorrow) //’

as a speech act of assertion followed by a greeting (or another assertion);

(3) João // vai pro Rio // até amanhã //
‘João // go to Rio // no later than tomorrow (*or* see you tomorrow) //’

as a recall, followed by an order and by an assertion (or a greeting);

(4) João vai pro Rio até amanhã //
‘João will go to Rio no later than tomorrow //’ (assertion or many different illocutions, such as - according to Portuguese syntactic structure - question, expression of surprise, expression of disbelief, etc.)

as just one speech act of assertion;

[2] In the examples of this paper, we mark non-terminal prosodic breaks with a one-slash sign and terminal prosodic breaks with a double-slash sign. The non-terminal break marks the boundary of a tone unit; the terminal one marks the boundary of an utterance.

(5) João / vai pro Rio até amanhã //
'João / will go to Rio no later than tomorrow //'

in which the non-terminal break after João signals that what is at its left and what is at its right must be considered two different information units of the same utterance, independently of the illocution marked in one of the units.

It is easy to imagine different segmentations with different speech acts, even in a very simple stretch like this one. What we can argue based on these examples is that the same lexical material in the same sequential order can convey a different number of utterances and different meanings, informational or morphosyntactic interpretations. For example, in (2) and (4) vai is a third person singular, present tense form and João is its subject, but in (3) the same lexical content (vai) is the second person, imperative form and João is not a subject. Similarly, if in (4) até amanhã can be analyzed as an adjunction to the VP, this is not possible in (3). In (5), the main difference compared to (4) is that João is isolated in a different prosodic unit, but does not form a different utterance as in (3). Depending on the prosodic realization, (5) can be analyzed as a Topic followed by an illocution or by an illocution followed by a different information unit.

What allows for the same sequence of words to be analyzed differently is only the prosodic ways in which they are performed. In these cases, prosody can convey segmentation (1, 2 or 3 utterances), the speech act types and the information structure inside the utterance (as in 5). Studies based on corpora demonstrate that what allows the perception of boundaries is the prosodic break, frequently without any pause, and that, on the other hand, it is very common to have even long pauses inside the same utterance[3].

We can define an utterance as the smallest stretch of speech with pragmatic and prosodic autonomy (Cresti 2000). This means that the utterance is the smallest stretch interpretable in autonomy, i.e. a speech act (Austin 1962). The utterance can be considered as the unit of reference for spontaneous speech. The utterance can be simple, when it features only the illocutionary unit, which is necessary to give interpretability and autonomy to the stretch, or it can be compound, when, before or after the illocutions, more non-illocutionary informational units are present[4]. Non-illocutionary information units can be of

[3] For a more in-depth treatment of speech segmentation see, among others, Cresti & Gramigni (2004) and Raso *et al.* (2015), where it is statistically demonstrated that no duration of pauses correlates with utterance or tone unit boundary.

[4] For a more in-depth presentation of this theoretical framework for speech segmentation and tagging (*Language into Act Theory*), which is based on spontaneous speech corpora analysis,

several types. For our purposes, we can concisely say that there are just two types of informational units: textual units and dialogic units. The formal counterpart of each informational unit is the tone unit.

Textual units make up the semantic text of the utterance. The illocutionary and some non-illocutionary units (topic, parenthetic, appendix and locutive introducer) have this kind of property. Dialogic Units (DU), which are the main focus of this paper, do not constitute the text of the utterance; they govern the interaction between speaker and addressee, fulfilling different functions. What we propose is that DMs can be better understood if they are inserted in an informational framework, corresponding to what we have called Dialogic Unit. Therefore, from now on, DM and DU should be taken, in our framework, as two different names for the same object.

An example taken from the corpus and its corresponding audio files can clarify the segmentation criteria and what we mean by tone unit and information unit.

1-6 (6) *SAB: c'ha la terza moglie / ora // s'è sposato tre volte // sì / infatti s'è detto / forse ci s'ha qualche speranza hhh // (ifamdl09, 102-104[5])
'he is married to his third wife / now // he married three times // yes / in fact we said / maybe we have some hope //'

see Cresti (2000) and Moneglia & Raso (2014). See also the studies available for download at < http://www.c-oral-brasil.org/> and < http://lablita.dit.unifi.it/>.

[5] The abbreviations for the source of the examples give the following information: the first letter (*i* or *b*) means *Italian* or *Brazilian*; *fam* or *pub* specify the corpus context, if private-familiar or public; *dl*, *cv* and *mn* refer to the dialogic, conversational or monologic section of the corpus. Then, the number of the text for each section is informed, followed, in square brackets, by the utterance(s) number(s). The three starred letters before the text indicate the speaker.

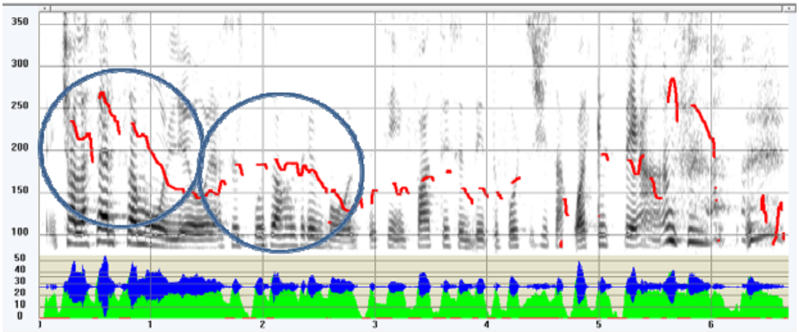

Figure 1. Sequence of utterances in ex. (6). The profiles of the first two are circled. No pause at the boundaries among the three utterances.

Audio 1 contains the whole sequence. Audios 2, 3 and 4 allow one to listen to each utterance in isolation. Audio 5 shows that the illocutionary tone unit of the first utterance is sufficient to guarantee the pragmatic and prosodic interpretability, whereas audio 6 shows that the two non-illocutionary units of the third utterance do not guarantee the interpretability without the illocution.

Of course, in order to know how trustable a segmentation is, we need a statistical measure of the inter-rater agreement among more segmentations. Both C-ORAL-ROM and C-ORAL-BRASIL reached an excellent score in the Kappa test[6].

3. How to predict DMs/DUs

Considering the framework summarized in 2, we can expect that a lexical item (or a very small set of lexical items) can appear in three possible contexts:

- it can fulfill an illocutionary function, therefore being pragmatically and prosodically autonomous, allowing its interpretability in isolation;
- it can appear inside a tone/information unit, being therefore semantically and syntactically compositional with the rest of unit's content, but not interpetable in isolation;
- it can fulfill a tone/information unit, though not performing an illocutionary function and therefore not being autonomous and interpretable in isolation.

[6] For the Kappa test, see Fleiss (1971); for the C-ORAL-ROM validation, see Moneglia *et al.* (2005); for the C-ORAL-BRASIL validation, see Mello *et al.* (2012), where the whole process of segmentation is also described.

We present two examples for each possibility, using the lexical items *vedi* in Italian and *não* in Brazilian Portuguese:

7 8 (7) *FER: *vedi* // la metti dentro / fa finta di pigliarla / e poi la ributta fuori // *vedi* // non la vuole // (ifamcv15, 42-45)
'you see // you put it inside / it pretends to take it / and then it throws it away // you see // it doesn't want it //'

In example (7), the lexical item *vedi*, repeated twice, is illocutionary. It can be interpreted as a speech act. Audio 7 presents the whole sequence, whereas audio 8 allows perceiving the illocutionary status of *vedi*.

9 10 (8) *SAB: poi / in piedi / hai visto / anche se il palco è un po' rialzato / però / se ti viene uno davanti alto / non *vedi* nulla // (ifamdl09, 40)
'then / standing / you've seen it / even if the stage is a little raised / however / if someone tall is before you / you don't see anything //'

Example (8) shows a case in which *vedi* is inside a textual unit, being compositional with the items before and after it. Listening to audio 9, the entire utterance can be appreciated, observing that no break separate *vedi* from the items with which it is compositional; listening to audio 10 it is possible to appreciate that the item *vedi* cannot be interpreted in isolation.

11 12 (9) *LID: no / poi / *vedi* / succede questo // (ifamdl02, 611)
'no / then / look / this is what happens //'

In example (9), the item *vedi* is isolated in a tone unit, but it is not illocutionary, not allowing an interpretation in isolation as in (7). At the same time, it is not compositional neither with what comes before it nor with what follows. Audios 11 and 12 present, respectively, the whole utterance and only the lexical item *vedi*.

Only in example (9) the lexical item *vedi* functions as a DU/DM. It is isolated in a tone unit, it does not perform an illocution (i.e. it is not interpretable in isolation), and it is not syntactically and semantically compositional with the rest of the utterance.

13 14 (10) *PAU: *não* // tá dando a altura daquele que a Isa marcou lá / né // (bpubdl01, 14)
'no // it reaches the height that Isa marked there / doesn't it //'

Example (10) shows that the item *não* performs an illocution and fulfills a simple utterance, being interpretable in isolation. If it were compositional with what follows it, the meaning would be *it does not reach the height that Isa marked there / isn't it.* Audio 13 allows listening to the whole sequence, while audio 14 allows appreciating the full interpretability of the lexical item *não* in isolation.

(11) *PAU: ah / *não* acaba não / acaba // (bpubdl01, 119) 15
‘ah / it doesn't end / does it //’ 16

Example (11) shows that the same item *não* can be compositional with what follows, marking the negative value of the verb (*acaba*). In the example we have one of the three strategies for negation in BP, with pre-verbal and post-verbal negation. Through audio 15, the whole utterance can be heard, whereas audio 16 shows only the preverbal *não*.

(12) *PAU: ah / *não* / ea disse que é pa ficar / por algum tempo // 17
(bpubdl01, 197) 18
‘ah / no / she said it must stay / for a while //’ 19

In (12), the same lexical item is neither illocutionary nor compositional. In this case, as the audios 17 and 18 show, *não* does not negate any semantic content and it is not interpretable in isolation. This allows us to conclude that it is used as a DM/DU. Notice that *ah*, which can be heard on audio 19, is also a DM.

Through examples 7-12 we wanted to show that a lexical item can have three different status: an illocutionary one, a compositional one and a DM/DU one; and that what allow us to predict which is its function is the combination of prosodic and syntactic cues: if the lexical item is isolated in a tone unit, it is not compositional; in this case it is illocutionary if it can be interpreted in isolation (it performs an utterance by itself and shows prosodic and pragmatic interpretability), and it is a DM/DU if it cannot be interpreted in isolation and needs to be interpreted along with the illocutionary part of the utterance.

There is just one situation in which it is possible that a tone unit is fulfilled by a lexical item that is still compositional with the items of previous or following units; in this case the units are called *scanned units*. The tone unit is considered to be the formal counterpart of an information unit, as we have already seen in 2. Although, sometimes it happens that, because of emphasis, size of the information unit (therefore, for articulatory reasons) or, more likely, some problem in the execution of an information pattern, parts of the same information unit are separated in different tone units. Example (12), repeated here, features a case of scanned units:

(12’) *PAU: ah / não / *ea disse que é pa ficar* / *por algum tempo* //
‘ah / no / she said it must stay / for a while //’

Here, the sequence *ea disse que é pa ficar por algum tempo* fills up two tone units that together form one information unit, namely an illocutionary one. When an information unit is performed in more than one tone unit, the prosodic profile that conveys the informational function is on the last tone unit. Two main features allow recognizing when a tone unit is a scanned unit and pertains to a wider informational program: the lack of functional prosodic profile on it and the syntactic compositionality with the following unit(s)[7]; in fact, only textual units can be scanned. This does not create any problem for the identification of DM/DU, except in a specific case that will be analyzed at the end of the next section.

There is still one point to be clarified in this section. Someone could think that interjections could be treated automatically as DM/DU, since they are always non-compositional. Nevertheless, even interjections should be submitted to prosodic analysis, as they can easily fulfill illocutionary units. In example (12) the interjection *ah* is not interpretable in isolation, as audio 19 allows perceiving, and is therefore a DM; but in the following example it is very clear that the same interjection is an illocution, as can be heard through audios 20 and 21:

20
21

(13) *FLA: esses dias / a gente tava procurando toalha / pa cobrir o /
carneirinho // que a gente disseca // <que ele> [/1] ele +
*REN: <o quê> //
*FLA: na aula de anatomia //
*REN: <*ah*> // (bfamdl01, 85-90)
‘these days / we were looking for a towel / to cover the / little lamb // that we dissect // that it / it +
what //
in the anatomy class //
ah //’

[7] For the relation between syntactic compositionality and information structure, see Cresti (2014).

4. Prosody marks DM/DU's functions

4.1 Our starting point

Previous studies have brought forth a proposal of six different DUs, each one with a specific function (Cresti 2000; Frosali 2008; Raso 2014). According to these studies, the different DUs in speech and their characteristics are the following:

- *Incipit* (INP). Its function is to take the turn or to mark affective contrast with the previous utterance. Its prosodic features are: rising, falling or rising-falling profile, very short duration and very high intensity. Its distribution is the first position of the utterance or of a sub-pattern of major terminated units called *Stanzas* (Cresti 2009).
- *Conative* (CNT). Its function is to push the interlocutor to do or quit doing something. Its prosodic features are: falling profile, short duration and high intensity. Its distribution is free.
- *Allocutive* (ALL). Its function is to identify the interlocutor and more frequently to mark social cohesion with him. Its prosodic features are: falling or flat profile, short duration and low intensity. Its distribution is free.
- *Phatic* (PHA). Its function is to keep the channel open. Its prosodic features are: falling or flat profile, short duration and low intensity. Its distribution is free.
- *Expressive* (EXP). Its functions are two: to take the turn without contrast and to support the illocution. Its prosodic features seem to be variable: different profiles with medium duration and intensity. Its distribution is free but mainly initial, maybe because of the different functions.
- *Discourse Connector* (DCT). Its function is to connect utterances or sub-patterns of Stanzas marking continuity. Its prosodic features are: different profiles, very long duration and high intensity. Its distribution is always at the beginning of the utterance or sub-pattern of Stanza.

What is important to emphasize in these studies is that, while prosody seems to present regularities that reflect functional differences, no regularity could be assigned to lexical items. Raso (2014) presents many lexical tables showing how rich can be, both in Italian and BP, the lexical variability for functions that seem to be strongly coherent if prosody is taken into account as their main formal counterpart. When lexicon is considered, we can find some correlations between lexicon and function, but it is abundantly clear that the observation of lexicon as the main formal counterpart of function would lead to a great confusion. For instance, while it is very difficult to find lexical correlations for INP, CNT and PHA, we can say that DCTs correlate with prepositions and conjunctions, which still means many lexical items, and that ALLs are lexically fulfilled by proper names, title and epithets (like *doctor*, *dear*, *honey* and many others). But what is clear is that the same lexical item can fulfill different functions. For instance, *because* can fulfill at least the functions of INP and DCT, and a proper name or titles and epithets can fulfill either ALLs or CNTs. What really marks the function is not the lexical item but its prosodic features.

Starting from these descriptions, we decided to statistically verify the possibility of distinguishing the groups only using prosodic features. We decided not to include the Expressives, since their prosodic description is vague, which probably conceals more than one DM/DU under its categorization. Expressive should be better studied, looking for more regularities between functions and different prosodic profiles.

Since we are analyzing spontaneous speech, where different voices are involved in different contexts and with a very high degree of variability in terms of linguistic and non linguistic conditions, it was necessary to adopt a term of comparison in order to establish how to judge different measurements, like what we called long or short duration and high or low intensity. We maintained the proposal of Raso (2014) to consider as term of comparison the measurements of the illocutionary unit of the same utterance in which the DM/DU is found. This does not completely eliminate the variability of the term of comparison, since different illocutions or different attitudes of the same illocution can change its value, but reduces variability very much, since the illocution is the only unit present in all the utterances. Moreover, we assume that prosodic parameters of DUs maintain the same tendency in relation to their illocution. For *Stanzas* (see Cresti 2009), where more illocutions can be found, we considered the one which appeared to be the real reference for the DM/DU under examination. *Stanzas* are organized in sub-patterns; therefore, it is usually simple to understand, in distributional terms, which one is the illocution of reference for a given DU.

4.2 Our first step

At this point, we made a first revision of the informational tagging of the BP minicorpus texts. The informational tagging had been carried for all units in the whole minicorpus, which necessarily causes some errors. We were aware that the risk of tagging errors was much higher in DUs, for the following reasons: DUs are smaller and therefore more sensible to the acoustic quality of the recording and of the single utterance; the fact that the semantic content of the DU cannot be considered as an important feature for its tagging implies that their prosodic features have an almost complete role in conveying their function. Therefore, even after the revision, we still knew that, without a better prosodic description (which is the main object of this paper), it would be hard to have a tagging without a high number of errors.

After the tagging revision, we selected all the DMs/DUs from the BP minicorpus. Our goal was to know whether some statistical analyses could give us at least indications to improve the description, and to verify if it was possible to clearly distinguish the five groups. Necessarily, we had to discard the cases in which overlapping or bad acoustic quality did not allow prosodic analysis. After this, DM/DUs and their corresponding illocutionary units were segmented and phonetically transcribed into syllables; and voiceless obstruents were ignored when it was impossible to identify their boundaries.

The software Praat (Boersma & Weenink 2014) was used for acoustic analyses. The *SG_Detector* script (Barbosa 2006) was used in order to normalize the duration of each syllable, taking into account the effect of the intrinsic duration for different segments. After previously setting the pitch range, the following measurements were taken using Praat script: raw duration (s), mean syllable duration (s), mean intensity (dB), speech rate (syllable/s), f0 variation rate (Hz/ms), mean f0 (Hz), minimum f0 (Hz), maximum f0 (Hz). Silent pauses are not taken into account by the script, which means that only speech was measured. As just said, because of the variability of contexts, it was important to take a unit as reference. So, the measurements were transformed into proportion of the difference (PD) in relation to their respective illocutionary unit (COM):

$$PD = ((DM- COM) / COM)*100$$

Preliminary analyses were conducted with linear mixed models (Baayen 2008) using the software *R* (R Core Team 2013). Random effect variables were audio, text and speaker, and fixed effect variables were previously gender and DU, along with other important variables that were not significantly correlated to each other. This statistical procedure was carried out just in order to point out which variables are more important to distinguish the DUs. The more general assumptions are verified through visual inspection. In total we analyzed the following number of DUs:

Table 1. Amount of data analyzed by unit

ALL	CNT	DCT	INP	PHA
33	14	42	38	33

Four variables are shown to be almost enough to identify algorithmically each DU: mean f0, mean intensity, normalized duration and articulation rate.

As for f0, the more significant model was built with the following fixed effect variables: DU, f0 variation rate, speech rate, intensity and gender. Speech rate (p=0.08) and f0 variation rate (p=0.058) turned out to be tendencies, but they were not significant. Gender was not significant (p=0.22), but it is important to maintain the linearity of the model, since differences between genders generate a positive correlation between residuals and fitted values. Paired-comparisons, by means of Tukey's test[8], showed that the significant effect of DU on mean f0 is due to the significant difference between INP and ALL, and between INP and PHA. INP is the DU with higher f0. The difference between INP and DCT (p=0.15), and between INP and CNT (p=0.11), with INP having higher mean f0, could perhaps be confirmed with more data. The same can be said with respect to the difference between DCT and ALL (p=0.15), with DCT having higher mean f0. No obvious pattern was found on the residual plot. The model could account for 40% of the data variability.

As for intensity, we built a final model with the following fixed effect variables: DU, mean f0, normalized duration, speech rate and number of syllables. A slightly linear pattern was found in the residuals, but it can be due to scale issues or to other variables with a small effect on them (for example, since intensity was correlated with both minimum f0 and maximum f0 in our data, we chose mean f0, which has a more global information). Assuming that this pattern does not strongly violate the linearity of the model, we can say that the significant effect of DU on mean intensity is due to the difference between CNT and ALL (p=0.01), INP and ALL (p=0.001), CNT and PHA (p<0.001), DCT and PHA (p=0.007), and INP and PHA (p=0.001), being in all these comparisons the former higher in intensity than the latter. A tendency for INP to be more intense than DCT was also found (p=0.08). This model accounts for 83% of the data variability. The number of syllables edges out the significance (p=0.051), which is expected since the larger the unit the more variable this parameter can be, because of the presence of more unstressed syllables, which are less intense than the stressed ones.

[8] All the paired-comparisons reported here were carried out by means of Tukey's test (analysis through multcomp package: Hothorn *et al.* 2008).

Normalized duration showed us just a strong tendency for DCT to have longer duration than INP (p=0.06). The final model was built with DU and speech rate as fixed effect variables. The lack of significant differences in duration among DUs may be due to the fact that those differences are not greater enough to overtake the effect of the final lengthening that they exhibit. This hypothesis is reinforced by the fact that if we insert in the model the position in the utterance as a predictor variable, the model shows the difference between DCT and INP as significant (p=0.04). This model accounted for 65% of the data variability. The model showed some limitation in predicting high values of normalized duration. The great variance of DCTs affects the model homoscedasticity. Later on, we will discuss the importance of the final position and also one possible reason for the great DCT variance. An alternative model with normalized duration varying between -20 and +20[9] was built. The same pattern was found, but this time DCT shows higher duration than CNT (p = 0.007) and INP (p=0.01). The difference between DCT and PHA is tangential (p= 0.06).

Finally, regarding speech rate, the model is less trustable. Linearity and homoscedasticity are somehow violated. Speech rate is correlated with several other variables and it is very difficult to measure its effect separately. The model does not appropriately account for extreme values. Adding mean f0 as a predictor variable reduces considerably the correlation between residuals and fitted values. However, mean f0 is not a significant factor and we do not see any reason to think that it has any effect on speech rate; on the contrary, we expect speech rate to have effects on mean f0, since it can generate phenomena such as truncation. The final model contains DU and intensity as fixed effect variables and accounts for 91% of data variability. The speech rate measurement was transformed into logarithmic scale. With the just mentioned caveat, the result is that all the DUs were significantly faster than DCT (DCT-ALL: p=0.03; DCT-CNT: p=0.001; DCT-INP: p= 0.005; DCT-PHA: p= <0.001), which is expected, since DCT is longer than the other DUs. We also found a tendency for PHA to be faster than ALL (p=0.08), which reaches the significance if we add number of syllables as a predictor (p=0.04).

At this point, we can propose a preliminary algorithm of identification:

[9] These values were chosen by means of inspection of the normalized duration in DCT. They exclude extreme values, while maintaining the DCT high variance.

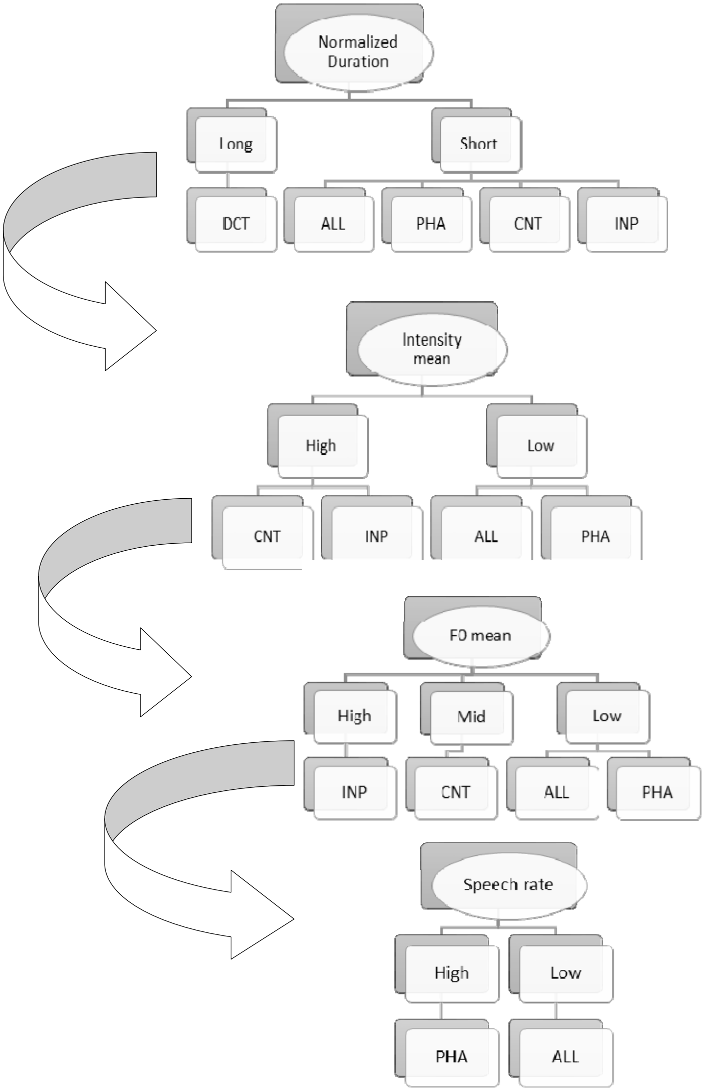

Figure 2. Step-by-step procedure for DUs identification

As previously mentioned, a few words about DUs' position in the utterance should be said. Some evident correlations between some DUs and their position were found; some DUs tend to appear exclusively or more frequently in a specific position: INP and DCT appear only in initial position; PHA and ALL have a tendency to avoid initial position; CNT seems distributionally free. In our small sample, these correlations are even stronger. This turns it difficult to quantify up to which extent the effects on the variables are due to the specific DU or to its position. However, it must be said that: i) we acknowledge that final position can lengthen the DU (final lengthening), and that it can decrease f0 (declination tendency) or decrease intensity (less air available); ii) DUs such as INP and DCT, and CNT and ALL often (or always) appear in the same position; therefore, the prosodic differences between them cannot be attributed to position. Also, CNT maintains its prosodic features when in initial position. As for ALL, we noticed that its profile changes according to the position: it is always falling in final position, but it seems to be flat and influenced by the surrounding units in medial position. We have not found yet, neither in BP nor in Italian, a unit with allocutive function in initial position[10]. Our hypothesis is that ALL has a flat profile, with interpolations between the lower f0 range of the ALL and the f0 range of the previous and/or following unit. Example 14, figure 3 and audio 22 show this kind of profile:

(14) *TON: é ocê / Onofre / que tá cagando assim // (bfamcv03, 247) 22
‘it is you / Onofre / who isn’t giving a shit //’

The models that we presented gave us many indications, even though they are not always conclusive. Moreover, these models do not take into account the intonation profiles of DUs, which, as we will see below, are one of the main features that allow us to differentiate among different DUs.

[10] Despite the fact that a first tagging of the Italian and Brazilian minicorpora does present some cases, those, in our opinion, do not satisfy the function of ALL and often seem to be CNT.

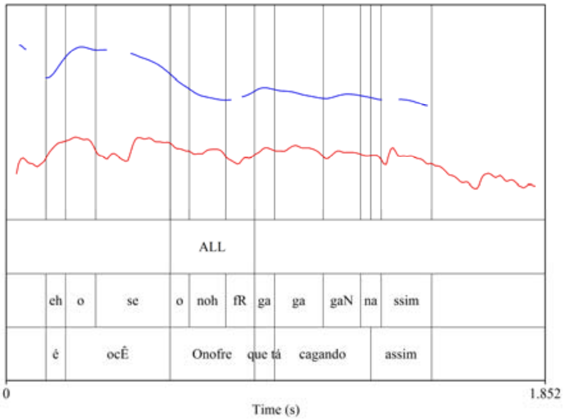

Figure 3[11]**.** Example of ALL in medial position

4.3 A more in-depth analysis

After these first results, we decided to concentrate on refining the prosodic description of the different units in order to:

a. make sure that the prosodic description could allow a trustable formal description of different functions;
b. reduce tagging errors to a minimum, before collecting more data for a final statistic test in future works.

Some aspects of previous descriptions were not convincing. In functional terms, besides Expressives, we were not convinced about Phatics. We also observed that units with high intensity in final position were frequently tagged as PHA. Therefore, we decided to exclude also PHAs from our analysis, since we thought that inconsistencies were too strong and needed a specific treatment.

Regarding prosody, we wanted to better understand mainly:

- why INP could be performed with three different forms;
- how could we better describe the prosodic differences between ALLs and CNTs, two units that were very convincing from the functional point of view but that could be easily confused using the prosodic parameters mentioned in previous studies.

[11] Here and in the following figures, we adopt broadly phonetic transcription system used in *SG_detector* (Barbosa 2006), with some adaptations.

We concentrated on a more in-depth prosodic analysis of INP, ALL, CNT and DCT, and we reached some interesting results that can provide better descriptions for these units.

4.3.1 *Incipit profiles*

Regarding intonation profile, INP must reach a very high f0 range in relation to the mean range in its illocutionary unit. That seems to be the only intonation feature necessary to convey the INP function along with other prosodic features such as short duration and high intensity. In our data inspection, correlations between INP intonation profile and different types of consonants and syllable structures are found. Thus, it is reasonable to think that the different profiles are due to the micromelodic effects of consonants on f0. Following this prediction, what would be expected, from the stressed syllable, is that: a) when an INP has more than one syllable or a diphthong, after reaching the high range, f0 would fall in order to perform other intonation patterns, which would hardly be higher than that of INP itself (falling profile); b) when an INP begins with a vowel, or a vowel preceded by a voiceless obstruent, the speaker is able to reach a high pitch immediately, either because it begins with a vocalic sound with no consonantal effects on it, or it begins with a voiceless obstruent whose effect is to increase f0, which contributes to reach the high range (flat profile); c) when an INP begins with a voiced obstruent followed by a vowel, the effect of lowering f0 due to this type of consonant generates a rising movement from the consonant to the vowel (rising profile). These predictions can be combined in order to account for more complex profiles, such as the rising-falling profile, which is due to the co-occurrence of b) and a); or a flat-falling profile, which can be due to the co-occurrence of c) and a). Notice that, since these are phonetic effects, the actual pronunciation and not the underlying phonologic representation must be taken into account. The following examples can clarify what we have described.

Examples 15, audio 23 and figure 4 show the case of disyllable [pu.ke], that, since it begins with voiceless consonant, is flat, as expected:

(15) *CEL: porque / vai ficar ruim no quatro // (bfamcv03, 276) 23
‘because / it will be bad for ball number four //’

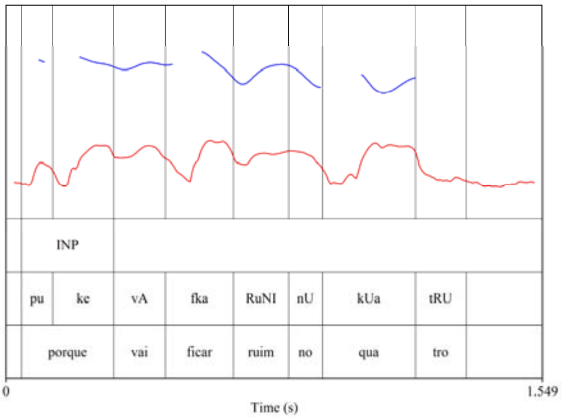

Figure 4. Example of INP with flat profile

Example 16, audio 24 and figure 5 show a case of an INP beginning with a voiced consonant followed by a diphthong. As it is expected, its profile is rising-falling:

24 (16) *BAO: não / mas é porque eu tô pensando assim // (bfamdl02, 197)
‘no / but it’s because I think this way //’

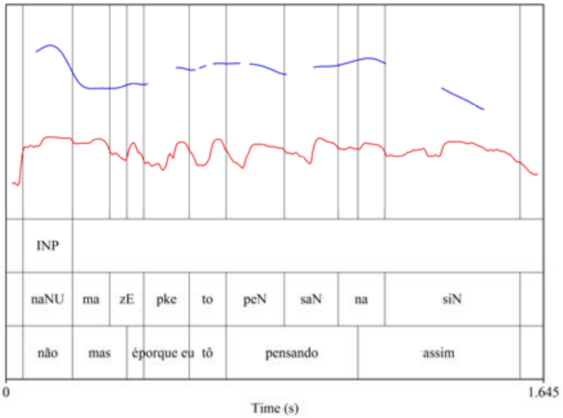

Figure 5. Example of rising-falling INP

Example 17, audio 25 and figure 6 show the case of a monosyllabic INP beginning with a voiced consonant. As it is expected, its profile is rising:

(17) *JOR: bom / eu fiquei nesse [/1] nessa empresa durante algum tempo / (bfammn06, 52) 25
‘well / I remained in this company for a while /’

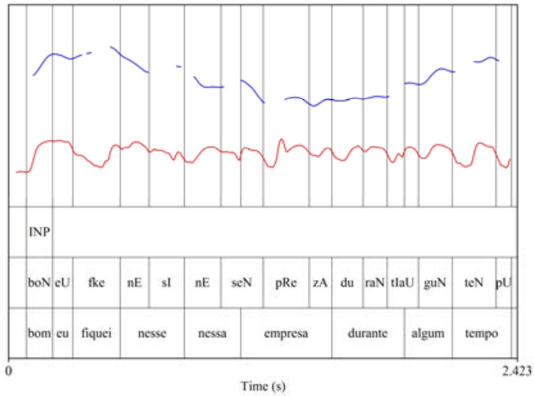

Figure 6. Example of rising INP

Example 18, audio 26 and figure 7 show the case of an INP beginning with an unvoiced consonant followed by a diphthong. As it is expected, its profile is falling:

(18) *ILA: poi / alla FLOG / cioè / non si sente mai nulla // (ifamcv06, 32) 26
‘then / at the FLOG / I mean / we never hear anything //’

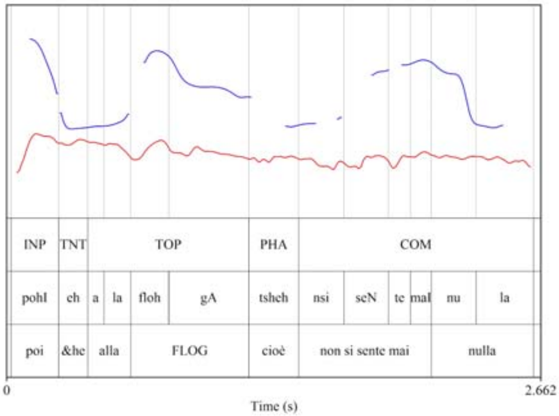

Figure 7[12]**.** First example of falling INP

[12] Notice that the low profile of the syllables [eh.a.la] is due to glottalization.

Example 19, audio 27 and figure 8 show the case of a disyllable INP [kũ.kue], stressed on the first syllable beginning with an unvoiced consonant. As it is expected, its profile is falling:

27 (19) *ELA: comunque / Massimo / io / la prima volta che l’ho vista / che anno era // (ifamcv01, 780)
‘anyway / Massimo / I / the first time I met her / what year was it //’

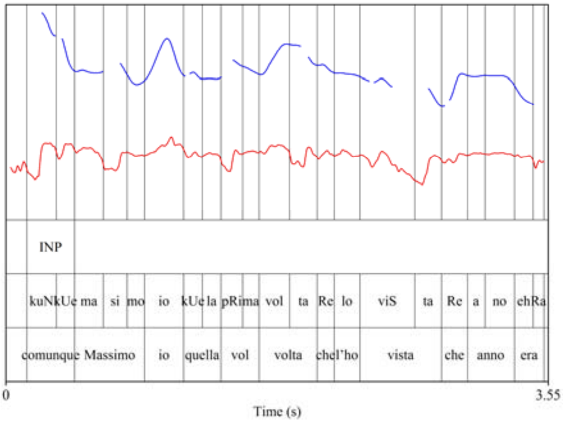

Figure 8. Second example of falling INP

4.3.2 *Allocutives (ALL) and conatives (CNT)*

ALL and CNT could sometimes be confused considering the prosodic description reported in 4.1. In fact, if the formal distinction is left only to measurements of duration and intensity, which are not always very clear, the decision can be left only to the subjective perception of the function. A careful analysis of many ALLs and CNTs allowed us to describe both forms more appropriately.

When the two units occur inside the utterance, as we said at the end of 4.2, their profile differs, since CNT maintains the falling profile while ALL shows a flat profile with interpolations. The confusion is possible when the unit appears in final position, which is a very common position, especially for ALL. In this case the first element that plays a distinctive role is the alignment of the f0 movement. While ALL presents a falling movement that starts from the beginning of the unit, CNT starts to fall from the stressed vowel. This feature easily allows distinguishing the two units when the stressed syllable is not the first one.

However, there are other features that can distinguish the two units and that become crucial when the lexical material begins with a stressed syllable. Besides the higher intensity of CNTs with respect to ALLs, other prosodic features help in distinguishing them: the type of movement; the phonetic realization of the lexical material; and some lexical correlations.

Even though the direction of the movement may seem the same (but in initial and internal positions it becomes clear that it is not), in final position there are differences that can help even when the stress is on the first syllable and when we cannot trust in the intensity; mainly, the slope of the CNT movement seems higher than that of ALL. We can therefore say that f0 variation rate is higher in CNTs than in ALLs. Moreover, the falling movement in CNT is maintained until the end of the unit, whereas ALLs usually feature a falling movement in the first part of the unit followed by a flat part at the end. Besides this, the phonetic realization of the CNT is clearer and more complete, whereas ALLs tend to be pronounced in a more centralized way. Lexical correlations can also help: there is lexical overlapping between the two units when they are fulfilled by proper names, titles or epithets, but only CNTs can be fulfilled by different lexical items, like *olha* "look", *espera* "wait", *aqui* "here", etc. in BP and the corresponding items in Italian (Raso 2014).

The following examples, without any lexical variation, can show the prosodic differences between these two units.

Example 20, figure 9 and audio 28 show a case of CNT stressed on the second syllable, while example 21, figure 10 and audio 29 show the same lexical item realized as ALL. This allows us to easily compare the prosodic differences between the two units.

(20)[13] *LUR: não / adorou / Lelena // 28
'no / he loved it / Lelena //'

[13] Examples 20 and 21 were extracted from a corpus that has not been published yet; therefore, no identification can be provided.

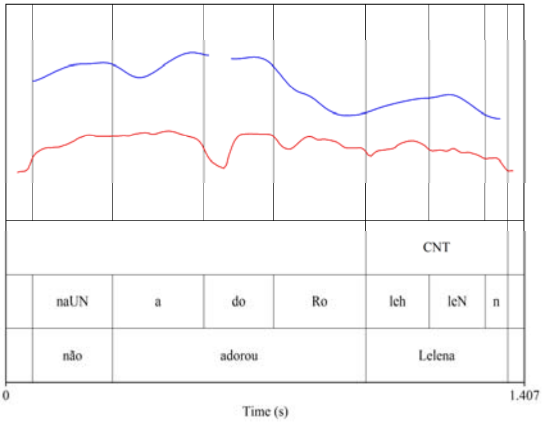

Figure 9. Example of CNT stressed on second syllable

29 (21) *LUR: oi / tudo bom / Lelena //
‘hello / everything OK / Lelena //’

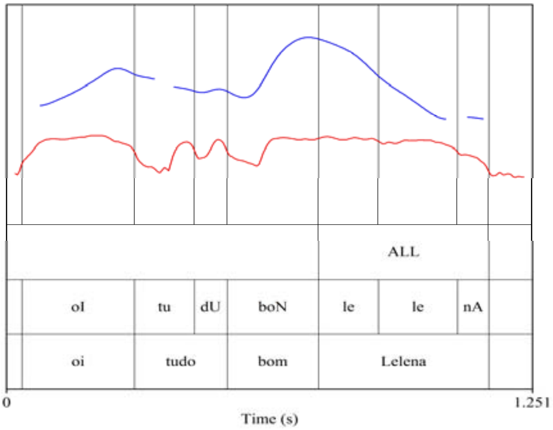

Figura 10. Example of ALL stressed on second syllable

Example 22, figure 11 and audio 30 show a case of CNT stressed on the first syllable, while example 23, figure 12 and audio 31 show the same lexical item realized as ALL. When the alignment of the falling movement cannot be used as the main feature that allows differentiating between CNT and ALL, we can better appreciate the other differences: intensity, slope of the movement and phonetic realization.

30

(22) *SIL: isso tudo é herança do tio dela / Kátia // (bfamdl04, 123)
‘all this is inheritance from her uncle / Kátia //’

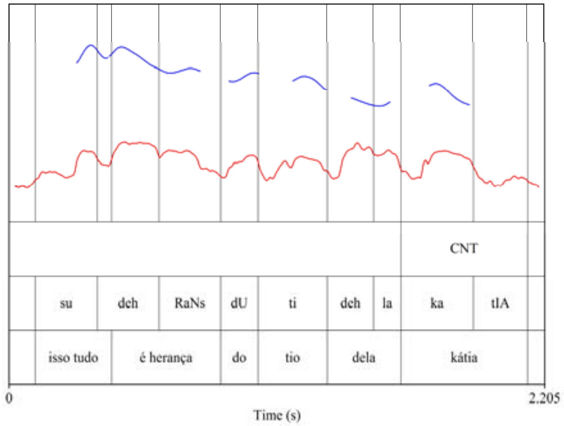

Figure 11. Example of CNT stressed on first syllable

31

(23) *LUR: cê leva aque’ cappelletti alí / Kátia // (bfamdl04, 191)
‘can you take those cappelletti over there / Kátia //’

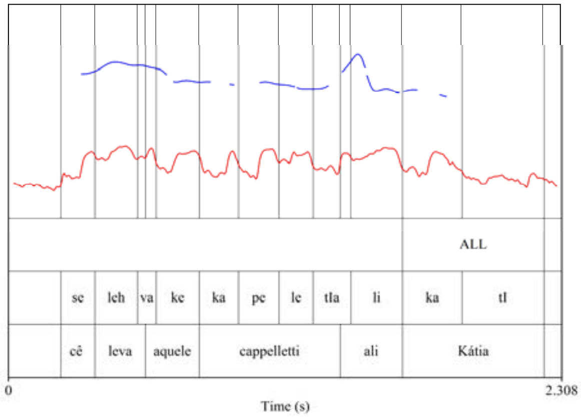

Figure 12. Example of ALL stressed on first syllable

4.3.2 *Discourse Connectors*

The prosodic description of DCTs by Raso (2014) has been kept. The main characteristic that allows the identification of this unit is its long duration, along with a high intensity. It also seems that its profile is not a condition for conveying the form, since it can be flat, slightly rising or rising-falling. F0 variation rate is usually low. There are correlations with some lexical classes, mainly conjunctions, but the lexical value by itself is not unambiguous. The new observations that emerged from our research deal more with the necessity of a better definition of this DM/DU and with the possible confusion of DCTs with scanned units. Also, in some cases, we found patterns in which items functionally candidate to be DCTs appear in internal position (initial of sub-pattern) and seem to have different prosodic characteristics, with shorter duration compared with when they are in initial position, but still longer than the mean duration of the illocutionary unit. Example 24, figure 13 and audio 32 show a DCT in these conditions.

32

(24) *GIL: é triste falar / mas / cê fala que é Futebol Arte / a galera começa a zoar / já // (bfamcv01, 136)
'it's sad to say it / but / you say it is Futebol Arte / people begin to mock / immediately //'

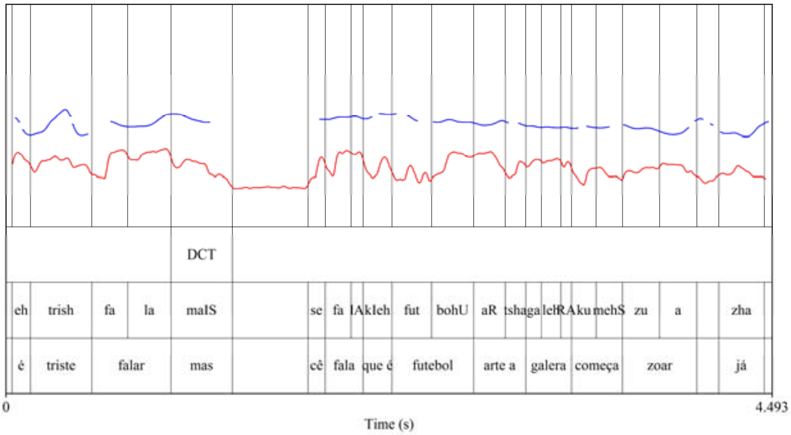

Figure 13. Example of DCT in internal position

DCT seems to have a different status from that of all the other DM/DUs. It is clearly more connected to the text and does not perform a function in governing the interaction with the interlocutor, at least not in the same sense as the other

DM/DUs do. This is the main reason for it to be the only DM/DU that appears more frequently in monologic than in dialogic texts (Raso 2014). Its function is to connect textual parts, and it is frequently present in *Stanzas*, connecting their sub-patterns, like in the two cases in example 25, or at the beginning of an utterance, whose content must be interpreted as marking continuation with the previous one, like in example 26. Audios 33 and 34 contain the two examples below:

(25) *ALO: eu nũ vou falar nome da cidade não / só pa nũ [/1] nũ compricar 33
a coisa / porque /=DCT= a dona Elvira tá viva ainda hhh / depois ea fica
sabendo disso / e pode querer acertar comigo / então /=DCT= melhor
ficar assim / do jeito que tá aí // (bfammn03, 11)
'I'll not mention the name of the town / just not to / not to complicate
things / because / dona Elvira is still alive hhh / then she finds it out / and
she may want to have it out with me / so / it's better to let things be / the
way they are now //'

(26) *DFL: <pruma família> religiosa / ter um filho padre era muito 34
importante //
*LUC: sim //
*DFL: e /=DCT= e sob o ponto de vista de [/1] de &cultu [/2] cultural /
ele ia adquirir muita cultura // (bfammn02, 95-97)
'for a religious family / to have a son who is a priest is very important //
yes //
and / and from the point of view / of cult / culture / he would acquire a
lot of culture //'

This leads to an ambiguous status of this unit, since it is positioned inside a complex pattern but in initial position of a sub-pattern. It is not clear whether this aspect can lead to prosodic effects. This different status is also confirmed by the fact that sometimes it is difficult to distinguish it from scanned units. This happens because DCT's semantic content seems frequently very important, and because its prosodic profile does not show specific characteristics. Besides these aspects, we have the impression that for a better description of DCT we would need a more in-depth syntactic analysis, since units that are candidates to be DCT frequently seem to have a peculiar syntactic and semantic scope. In example 27 (audio 35), *porque* "because" shows a shift from the propositional to an epistemic-pragmatic scope. In fact, it does not introduce a cause, but it means something like "I am saying this because...".

35 (27) *GIL: Cuecas / vão ver se vai querer participar /né // tão + *porque* / es tavam reclamando até // (bfamcv01, 176-178)
'Cuecas / let's see if they will participate / you know // so + because / they were complaining a lot //'

5. Conclusion: advances and limits in present knowledge of DM/DUs

The proposal of explaining DMs within an informational framework and of attributing to prosody an important formal role in conveying functions seems to be promising. The lexical perspectives did not seem to reach good results, neither in predicting DMs nor in explaining their specific functions. It makes more sense, we think, that the main cue that conveys functional values in speech needs a prosodic investigation, due to the crucial role of prosody and the fact that prosody is clearly less conventionalized than lexicon and less sensible to linguistic change over time (even a short span of time) and through different social groups.

From Cresti (2000) until this paper, passing through Frosali (2008) and Raso (2014), many aspects have been clarified. We think that four different functions with clear different formal features could be confirmed. In this paper, we were able to point out a reason for the different profiles of INP and establish what is actually pertinent to describe its form; we also provided a better description for ALLs and CNTs, making the prosodic distinction between them clearer. At the same time, we showed that DCT is more problematic than Raso (2014) proposed. It needs more investigation in order to better understand its function, which plays a more textually marked role, whereas the other DUs have a clearer interactional function. This is also confirmed by the fact that not always is it easy to distinguish DCT from a scanned unit, which is always part of a textual unit, together with the fact that DCT seems to have a stronger semantic value than the other DUs and that what seems to help to individualize it is its semantic and syntactic scope.

Regarding Expressives and Phatics, nothing can be said so far, except that they certainly need a specific study. It is not clear whether they are just two DUs or whether under these tags we can find more different units.

Acknowledgments

This research was possible thanks to CNPq and Fapemig fundings.

References

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

Baayen, R. 2008. *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Barbosa, P.A. 2006. *Incursões em torno do ritmo da fala*. São Paulo: Pontes.

Bazzanella, C., Bosco, C., Gili Fivela, B., Mieznikowski, J. & Brunozzi, F.T. 2008. Polifunzionalità dei segnali discorsivi, sviluppo conversazionale e ruolo dei tratti fonetici e fonologici. In M. Pettorino, A. Giannini, M. Vallone & R. Savy (eds), *La comunicazione parlata*, Vol. 2. Napoli: Liguori, 934-963.

Bolden, G. 2015. Discourse Markers. In K. Tracy, C. Ilia & T. Sandel (eds), *The International Encyclopedia of Language and Social Interaction*. Hoboken: John Wiley & Sons, 1-7.

Boersma, P. & Weenink, D. 2013. *Praat: doing phonetics by computer*. [Computer program] (accessed December 1, 2016).

Cresti, E. 2000. *Corpus di Italiano parlato*, Vol. 1. Firenze: Accademia della Crusca.

Cresti, E. 2009. La Stanza: un'unità di costruzione testuale del parlato. In A. Ferrari (ed.), *Sintassi storica e sincronica dell'italiano. Subordinazione, coordinazione e giustapposizione, Atti del X Congresso dela Società Internazionale di Linguistica Italiana (SILFI).* Firenze: Franco Cesati, 713-732.

Cresti, E. 2014. Syntacic properties of spontaneous speech in the Language into Act Theory. In T. Raso & H. Mello (eds), *Spoken corpora and linguistic studies*. Amsterdam: John Benjamins, 365-410.

Cresti, E. & Gramigni, P. 2004. Per una linguistica corpus based dell'italiano parlato: Le unità di riferimento. In F. Albano Leoni, F. Cutugno & M. Pettorino (eds), *Atti del Convegno Nazionale "Il Parlato Italiano"*. Napoli: D'Auria, 1-26.

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

Fischer, K. (ed.) 2006. *Approaches to Discourse Particles*. Leiden: Brill.

Fleiss, J.L. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76: 378-382.

Frosali, F. 2008. L'unità di informazione di ausilio dialogico: Valori percentuali, caratteri intonativi, lessicali e morfo-sintattici in un corpus di italiano parlato (C-ORAL-ROM). In E. Cresti (ed.), *Prospettive nello studio del lessico italiano*. Firenze: Firenze University Press, 417-424.

Hothorn, T., Bretz, F. & Westfall, P. 2008. Simultaneous Inference in General Parametric Models. *Biometrical Journal* 50(3): 346–363.

Mello, H. & Raso, T. 2011. Illocution, Modality, Attitude: different names for different categories. In H. Mello A. Panunzi & T. Raso (eds), *Pragmatics and Prosody: Illocution, Modality, Attitude*, *Information Patterning and Speech Annotation*. Firenze: Firenze University Press, 1-18.

Mello, H., Raso, T., Mittmann, M., Vale, H. & Côrtes, P. 2012. Transcrição e segmentação prosódica do corpus C-ORAL-BRASIL: critérios de implementação e validação. In T. Raso & H. Mello (eds), *C-ORAL-BRASIL I. Corpus de referência do português brasileiro falado informal.* Belo Horizonte: Universidade Federal de Minas Gerais, 125-176.

Mittmann, M. & Raso, T. 2011. The C-ORAL-BRASIL informationally tagged *minicorpus*. In H. Mello, A. Panunzi & T. Raso (eds), *Illocution, modality attitude, information patterning and speech annotation*. Firenze: Firenze University Press, 151-183.

Moneglia, M. 2005. The C-ORAL-ROM resource. In E. Cresti & M. Moneglia (eds), *C-ORAL-ROM: integrated reference* corpora *for spoken Romance languages.* Amsterdam: John Benjamins, 1-70.

Moneglia, M., Fabbri, M., Quazza, S., Panizza, A., Danieli, M., Garrido, J.M. & Swerts, M. 2005. C-ORAL-ROM prosodic boundaries for spontaneous speech analysis. In Y. Kawaguchi, S. Zaima & T. Takagaki (eds), *Spoken Language Corpus and Linguistic Informatics*. Amsterdam: John Benjamins, 257-276.

Moneglia, M. & Raso, T. 2014. Notes on Language into Act Theory. In T. Raso & H. Mello (eds), *Spoken corpora and linguistic studies*. Amsterdam: John Benjamins, 469-495.

Panunzi, A. & Mittmann, M. 2014. The IPIC resource and a cross-linguistic analysis of information structure in Italian and Brazilian Portuguese. In T. Raso & H. Mello (eds), *Spoken corpora and linguistic studies*. Amsterdam: John Benjamins, 129-151.

R Development Core Team, 2013. *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna: the R Foundation for Statistical Computing http://www.R-project.org/ (accessed December 1, 2016).

Raso, T. 2014. Prosodic Constraints for Discourse Markers. In T. Raso & H. Mello (eds), *Spoken corpora and linguistic studies*. Amsterdam: John Benjamins, 412-467.

Raso, T. & Mello, H. (eds) 2012. *C-ORAL-BRASIL I. Corpus de referência do português brasileiro falado informal.* Belo Horizonte: Universidade Federal de Minas Gerais.

Raso, T., Mittmann, M. & Oliveira, A.C. 2015. O papel da pausa na segmentação prosódica de corpora de fala. *Revista de Estudos da Linguagem* 17(2): 883-922.

Schourup, L. 1999. Discourse Markers. *Lingua* 107: 227-265.

Traugott, E.C. 2007. Discourse markers, modal particles, and contrastive analysis, synchronic and diachronic. *Catalan Journal of Linguistics* 6: 139-157.