STRUMENTI
PER LA DIDATTICA E LA RICERCA

– 120 –

UNIVERSITÀ DEGLI STUDI DI FIRENZE
DIPARTIMENTO DI ITALIANISTICA

BIBLIOTECA DIGITALE

Volumi pubblicati:

Moderna [diretta da Anna Dolfi]

1. *Giuseppe Dessí. Storia e catalogo di un archivio,* a cura di Agnese Landini, 2002.
2. *Le corrispondenze familiari nell'archivio Dessí*, a cura di Chiara Andrei, 2003.
3. Nives Trentini, *Lettere dalla Spagna. Sugli epistolari a Oreste Macrí*, 2004.
4. *Lettere a Ruggero Jacobbi. Regesto di un fondo inedito con un'appendice di lettere*, a cura di Francesca Bartolini, 2006.
5. *«L'Approdo». Copioni, lettere, indici*, a cura di Michela Baldini, Teresa Spignoli e del GRAP, sotto la direzione di Anna Dolfi, 2007 (CD-Rom allegato con gli indici della rivista e la schedatura completa di copioni e lettere).
6. Anna Dolfi, *Percorsi di macritica*, 2007 (CD-Rom allegato con il *Catalogo della Biblioteca di Oreste Macrí*).
7. *Ruggero Jacobbi alla radio*, a cura di Eleonora Pancani, 2007.
8. Ruggero Jacobbi, *Prose e racconti. Inediti e rari*, a cura di Silvia Fantacci, 2007.
9. Luciano Curreri, *La consegna dei testimoni tra letteratura e critica. A partire da Nerval, Valéry, Foscolo, d'Annunzio*, 2009.
10. Ruggero Jacobbi, *Faulkner ed Hemingway. Due nobel americani*, a cura di Nicola Turi, 2009.
11. Sandro Piazzesi, *Girolamo Borsieri. Un colto poligrafo del Seicento. Con un inedito «Il Salterio Affetti Spirituali»*, 2009.
12. *A Giuseppe Dessí. Lettere di amici e lettori. Con un'appendice di lettere inedite*, a cura di Francesca Nencioni, 2009.
13. Giuseppe Dessí, *Diari 1949-1951*, a cura di Franca Linari, 2009.
14. Giuseppe Dessí, *Diari 1952-1962*, trascrizione di Franca Linari, Introduzione e note di Francesca Nencioni, 2011.
15. Giuseppe Dessí, *Diari 1963-1977*, trascrizione di Franca Linari, Introduzione e note di Francesca Nencioni, 2011.

Linguistica [diretta da Massimo Moneglia]

1. *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*, edited by Massimo Moneglia and Alessandro Panunzi, 2010.
2. Alessandro Panunzi, *La variazione semantica del verbo* essere *nell'Italiano parlato*, 2010.
3. *Language, Cognition and Identity. Extensions of the endocentric/exocentric language typology*, edited by Emanuela Cresti and Iørn Korzen, 2010.
4. *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation*, edited by Heliana Mello, Alessandro Panunzi, Tommaso Raso, 2012.

Informatica e letteratura [diretta da Simone Magherini]

1. *BIL Bibliografia Informatizzata Leopardiana 1815-1999. Manuale d'uso vers. 1.0*, a cura di Simone Magherini, 2003.

Pragmatics and Prosody

Illocution, Modality, Attitude, Information Patterning and Speech Annotation

edited by
Heliana Mello
Alessandro Panunzi
Tommaso Raso

Firenze University Press
2011

Pragmatics and Prosody : Illocution, Modality, Attitude, Information Patterning and Speech Annotation / edited by Heliana Mello, Alessandro Panunzi, Tommaso Raso. – Firenze : Firenze University Press, 2011.
(Strumenti per la didattica e la ricerca ; 120)

http://digital.casalini.it/9788866550846

ISBN 978-88-6655-080-8 (print)
ISBN 978-88-6655-084-6 (online)

Progetto grafico di Alberto Pizarro Fernández
Immagine di copertina: Alessandro Geri Rustighi

Università degli Studi di Firenze
Firenze University Press
Borgo Albizi, 28, 50122 Firenze, Italy
http://www.fupress.com/
Printed in Italy

INDEX

Mello H., Panunzi A., Raso T. (eds), *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation*

INTRODUCTION

Most of the papers collected in this book resulted from presentations and discussions undertaken during the V Lablita Workshop that took place at the Federal University of Minas Gerais, Brazil, on August 23-25, 2011. The workshop was held in conjunction with the II Brazilian Seminar on Pragmatics and Prosody. The guiding themes for the joint event were illocution, modality, attitude, information patterning and speech annotation. Thus, all papers presented here are concerned with theoretical and methodological issues related to the study of speech. Among the papers in this volume, there are different theoretical orientations, which are mirrored through the methodological designs of studies pursued. However, all papers are based on the analysis of actual speech, be it from corpora or from experimental contexts trying to emulate natural speech. Prosody is the keyword that comes out from all the papers in this publication, which indicates the high standing of this category in relation to studies that are geared towards the understanding of major elements that are constitutive of the structuring of speech. This book also features a cluster of papers analyzing both Italian and Brazilian Portuguese, anchored on the Language Into Act Theory, proposed by Emanuela Cresti[1], and born out of the very LABLITA lab at Florence University.

Heliana Mello and Tommaso Raso propose the experimental investigation of three categories that often times are intertwined in the literatures: modality, attitude and illocution. Based on empirical findings, the authors suggest that introspective methodologies are inaccurate to evaluate the boundaries of such categories. Departing from a theoretical overview of modality, attitude an illocution and the observational results from empirical data, corroborated by the observations from a preliminary experiment, Mello & Raso suggest that modality should be considered a semantic category, the stance of the speaker toward the propositional content of his locution; while illocution should be applied to the action performed through an utterance, and the term attitude should point to the way this action is performed.

João Antônio de Moraes studies attitudinal meaning from a prosodic point of view, grouping emotions and social attitudes on the one hand and propositional attitudes on the other. Moraes advances the view that results of perceptive analysis,

[1] Cresti, E. 2000. *Corpus di Italiano Parlato*. Firenze: Accademia della Crusca.

Mello H., Panunzi A., Raso T. (eds), *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation* © 2011 Firenze University Press.

acoustic analysis and even F0 manipulation experiments with resynthesis reinforce the idea that there are two independent prosodic systems as proposed. Among his findings, obtained through experimental protocols, Moraes suggests that emotions and social attitudes do not conflict with speech acts or propositional attitudes since the phonological representation of a particular illocutionary act spoken with different emotional or social-attitudinal values would be the same: there are no localized, punctual F0 changes, but global modifications in the overall pattern (register and tonal span), not to be represented in phonological form. With propositional attitudes and speech acts, the changes are local (discrete), leading to distinct phonological analyses.

Emanuela Cresti, guided by empirical findings that motivated her Language Into Act Theory, advances the debate about pragmatic and semantic functions, showing that Focus and Comment are actually different categories; the former pertaining to the semantic level, therefore to a locutive act, while the latter is instantiated pragmatically and belongs in the illocutive domain. Through a detailed integration of spontaneous speech data and theoretical reasoning, Cresti demonstrates how Focus is a semantic property constrained within the boundaries of Topic and Comment information units.

Ida Tucci considers the differentiation between illocution and modality from an empirical stance, basing her analysis on data from the Italian corpus in the C-ORAL-ROM[2]. Tucci supports the view that although modal indexes may contribute to the illocutionary interpretation of an utterance, there is no direct correspondence between modal values and illocutionary forces. She further shows that each illocutionary type recorded in the corpus can express a range of modal values and the inverse is also the case, thereby modality and illocution are in a reciprocal distribution relationship. She the proposes that the scope of modality in spontaneous speech is constrained to an information unit which then differs from the scope of the illocutionary force which applies to the utterance. She concludes that modality is a semantic aspect of the locutive program, in which the speaker's stance towards his locutory expression is manifested, while the illocutionary force is a pragmatic effect through which the speaker manifests his attitude towards his interlocutor.

Sandra Madureira investigates speech expressivity and sets as foci of her paper the following: theoretical issues concerning speech expressivity and sound symbolism; the presentation of methodological procedures developed in the investigation of speech expressivity by the author and her research team, and the description of results of such methodological procedures for the analysis of expressivity in a speech. Spectrographic and perceptual analyses of the recording by two professional actors of the Sonnet of Fidelity were carried out and show that

[2] Cresti, E. & Moneglia, M. (eds) 2005. *C-ORAL-ROM: integrated reference corpora for Spoken Romance Languages*. Amsterdam/Philadelphia: John Benjamins.

voice quality settings play an important role in speech expressivity and should be considered in combination with intonation and duration patterns.

Plínio A. Barbosa posits the following question as his paper backbone: "what makes utterances sound prosodically distinct in different speakers, in different speaking styles and in different language varieties?" He then proposes that rhythm should be the prosodic domain through which the indicated variability could be best understood. He advocates coupled-oscillator theories as the methodological grounding for his study of rhythmic variability in Brazilian and European Portuguese and pursues his analysis from data collected through reading and storytelling tasks.

Alessandro Panunzi and Lorenzo Gregori present the DB-IPIC, an XML database for the study of the informational structure in spoken language. The linguistic data that have been inserted in the database derive from the Italian section of C-ORAL-ROM corpus. Transcripts have been implemented with the annotation of the informational structure, following the theoretical framework of the Language into Act Theory. The paper describes the procedure of annotation for the corpus. Starting from the collected data, the authors report some general measures regarding the referring units for the pragmatic analysis within the adopted theoretical framework: Utterances and Stanzas.

Maryualê Mittmann and Tommaso Raso devote their paper to the presentation of a mini-corpus extracted from the C-ORAL-BRASIL corpus[3] for spontaneous spoken Brazilian Portuguese. This corpus was tagged for informational structure and inserted in the DB-IPIC tool, therefore allowing for initial considerations about the information structure of Brazilian Portuguese in comparison to a similarly structured Italian mini-corpus. The authors find evidence differentiating the underlying processes of informational tagging from those of prosodic annotation, which leads to a better understanding of both the perceptual aspects related to the prosodic annotation and the cognitive aspects associated to informational tagging. The authors found that Brazilian Portuguese tends to use textual units much less often and to be more actional and less textual than Italian. At the same time, they observed that one particular textual unit, the locutive introducer, is much more used in Brazilian Portuguese.

Heliana Mello, Alessandro Panunzi & Tommaso Raso
December 2011

[3] Raso, T. & Mello, H. 2010. The C-ORAL-BRASIL corpus. In M. Moneglia & A. Panunzi (eds), *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*. Università degli studi di Firenze, 193-213.

ILLOCUTION, MODALITY, ATTITUDE: DIFFERENT NAMES FOR DIFFERENT CATEGORIES[1]

Heliana Mello, Tommaso Raso
Universidade Federal de Minas Gerais

1. Introduction

It is very likely that if given the three labels in the title of this paper, linguists from the same field of research would provide very different definitions for them. For some, at least two of those labels, but possibly even the three of them, could be lumped together and grouped within the same big category. What we are going to attempt to do in this paper, is not so much as to definitely define these three categories, but to argue that they indeed represent different sets of phenomena and must, therefore, be looked upon as such. In the process, we will hint at a tentative definition for each category bearing in mind that they are highly complex and demand in depth studying. We will also suggest that rather than just approaching these categories from an introspective theoretical point of view, it is high time they are seriously studied based both on empirical as well as experimental evidence.

We will discuss the labels illocution (or illocutionary act), modality and attitude and their various applications within the Linguistics literature, with special attention to the Pragmatics/Prosody interface. The fact that these categories can be found to correspond to the same or to overlapping concepts creates a lack of precision and clarity which would be desirable in technical terminology, and what is even more problematic, a confusion and imprecision of the very object of study. Thus, we aim at proposing specific scopes and characterizations for each of the labels and to start discussing possible criteria that would facilitate the identification of features that might lead to a repertoire of features for each of the concepts. In order to do so, we will first present some of the definitions and characterizations for illocution, modality and attitude found in the relevant literature and later will discuss ways to address the distinction for these three conceptual categories.

[1] This research was financed by CNPq and Fapemig.

Mello H., Panunzi A., Raso T. (eds), *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation* © 2011 Firenze University Press.

The label illocution, also referred to as illocutionary act, can be addressed among other possibilities as: (a) an act (1) for the performance of which one must make it clear to some other person that the act is performed (Austin speaks of the "securing of uptake"), and (2) the performance of which involves the production of what Austin calls "conventional consequences" as, e.g., rights, commitments, or obligations (Austin 1975: 116f., 121, 139); (b) an attempt to communicate in the expressing of an attitude (Bach & Harnish 1979); (c) the act of meaning something (Schiffer 1992: 103).

John Searle (1969) claims that the illocutionary act is «the minimal complete unit of human linguistic communication. Whenever we talk or write to each other, we are performing illocutionary acts». Searle posits five illocutionary points:

1. Assertives: statements that may be judged true or false because they purport to describe a state of affairs in the world;
2. Directives: statements that attempt to make the auditor´s actions fit the propositional content;
3. Commissives: statements which commit the speaker to a course of action as described by the propositional content;
4. Expressives: statements that express the "sincerity condition of the speech act";
5. Declaratives: statements that attempt to change the world by "representing it as having been changed".

Searle focuses primarily on the idea of a performative predicate which defines the act in his logic-lexicalizing approach, even when it has to be inferred and is not explicitly present in the act. On the other hand, Searle overlaps modality and illocution when he posits that to assert X and to assert I think that X stand for different acts. Additionally Searle admits indirect acts in his proposal.

Cresti (2001), pairing with Austin's ideas, asserts that the illocution co-occurs with the locutory act and functions as the affective engine of the linguistic act. It is related to the interpersonal dynamics of rapport, therefore the attitude towards the interlocutor – the Modus towards the Partner. Cresti (2000a; 2000b) proposes five illocutive classes: refusal, assertion, direction, expression and rite; these are subdivided into several possible subclasses. She individuates each act type based mostly, but not only, on prosodic criteria. Prosody would be the necessary (and at times sufficient) criterion to define an illocutionary act. Cresti, differently from Searle, separates illocution and modality.

The concept of illocution and the discussion about illocutionary force at times conflates with that of mood. Such is the case in Green (2009) who says that

> Mood together with content underdetermine force. On the other hand, it is a plausible hypothesis that grammatical mood is one of the devices we use, together

> with contextual clues, intonation and so on, to indicate the force with which we are expressing a given content.

According to Meyer (1997: 23) it is only fairly recently that the notions illocution and propositional attitude (modality) have been separated in the linguistics literature. According to him, authors such as Ungeheuer (1972) considered the two as the same category. On the other hand, he points out that Lüdtke (1980) proposes that modality and illocution are distinct, however, are connected as just being different kinds of propositional attitude. In his words, subjective modality shares with illocution the property of regarding the speaker's attitude or assessment concerning his proposition. The basic difference between them is that the latter concerns the relation between the speaker and the hearer – thus the goals and operations of communication –, whereas the former does not.

The label modality would encompass the following proposals: «the essence of "modality" consists in the relativization of the validity of sentence meanings to a set of possible worlds» (Keifer 1994: 2515a); from a speaker's-evaluation approach, modality is «the speaker's cognitive, emotive, or volitive attitude toward a state of affairs» (Keifer 1994: 2516a), his "commitment or detachment" , his "envisaging several possible courses of events" or his "considering of things being otherwise" (Keifer 1994: 2516b).

For Ruthrof (1991) modality is «the structurable field of the manners of speaking underlying all utterances» (this he also calls covert or inferential modality). Bybee & Fleischman (1995: 2) state that «Modality… is the semantic domain pertaining to elements of meaning that languages express. It covers a broad range of semantic nuances - jussive, desiderative, intentive, hypothetical, potential, obligative, dubitative, hortatory, exclamative, etc. - whose common denominator is the addition of a supplement or overlay of meaning to the most neutral semantic value of the proposition of an utterance, namely factual and declarative». Schneider (1999: 13) and Bybee (1985) point out that modality consists of (i) speech acts (orders and wishes, i.e. deontic modality), and (ii) attitudes to truth-content of the sentence (i.e. epistemic modality). Karkkainem (1987) claims that modality and illocutionary force are very similar since both express the speaker's attitude or opinion, therefore carrying the communicative purpose in the accomplishment of a speech act. Cresti (2001), following Bally (1950), asserts that modality expresses the speaker's attitude (modus) towards the content of an utterance, i.e., the referential or cognitive content (dictum). The major modal categories would be alethic, epistemic and deontic.

There are several proposed modal typologies which vary from the tripartite option followed by Cresti (2001) to Mindt's 17 modal meanings: (i) possibility/high probability, (ii) certainty/prediction, (iii) ability, (iv) hypothetical event/result, (v) habit, (vi) inference/deduction, (vii) obligation, (viii) advisability/desirability, (ix)

volition/intention, (x) intention, (xi) politeness/downtoning, (xii) consent, (xiii) state in the past, (xiv) permission, (xv) courage, (xvi) regulation/prescription, (xvii) disrespect/insolence (Mindt 1998: 45).

The label attitude, per se, is less discussed in the literature. Attitude is usually mentioned as an attribute inherent to illocution as well as to modality. However, some authors have explicitly mentioned attitude as a category on its own. According to Local (2005) «Attitude is widely acknowledged as making an important contribution to the meanings which can be attributed to utterances. Attitude is used as a cover term for constructs which have been referred to elsewhere as "attitude", "emotion", "affect" and "stance"». Local mentions that «in intonation studies there is a continuing tradition of employing lay attitudinal categories (e.g. "challenging", "surprised", "sad", "involved", "uncertain") in trying to account for the distribution and meaning of intonation contours (Cruttenden 1997; Schubiger 1958; Pierrehumbert & Hirschberg 1990; Ladd 1986)». He goes on to say that «within pragmatics, too, claims about particular pragmatic practices and stylistic effects (e.g. epistemic markers, facticity, irony, politeness, reported speech, sarcasm) and the intended force of utterances are routinely linked to speaker attitude (Mey 1993; Sperber & Wilson 1986; Leech 1983; Blakemore 1992)».

Additionally, attitude is related to a speaker's expression of social affects, voluntarily controlled by the speaker (Moraes et al. 2010). According to Moraes et al. (2010) there are attitudes that affect the propositional content of an utterance (irony, incredulity, obviousness, surprise, etc) and others that are connected to the social relationship established between interactants in a communication event (politeness, arrogance, authority, irritation, etc).

Attitude is at times conflated with emotion (Mozziconacci 2001) and therefore might be categorized as such into 48 different types according to the HUMAINE Emotion Annotation and Representation Language, which covers politeness, anger, courage, pride, serenity, empathy, happiness, among many others. In other views, attitude would refer to categories such as declarative, question, exclamation, incredulous question, suspicious irony and obviousness, therefore lumping illocutionary types with emotional types (Bailly & Holm 2002). In a crosslinguistic experimental setting, Shochi, Albergé & Rilliard (2006) study misperception of attitudes across Japanese and French. The authors studied 12 attitudes and, similarly to the assertions made above, there is the grouping together of categories that could easily be claimed to be related to either modality or illocution by other authors. Their list is: doubt-incredulity, evidence, exclamation of surprise, authority, irritation, arrogance-impoliteness, sincerity-politeness, admiration, kyoshuku, simple-politeness, declaration and interrogation.

As pointed out above, the definitions for illocution, modality and attitude vary and at times mix. In order to try to separate the domains of application of each of these concepts, we propose that they be established as instances of different

phenomena which apply at different levels of the communicative act, and can, in principle be compositional. This would need to be tested experimentally in order to be checked. Following and expanding Cresti (2001), we suggest a rationale that allocates modality to a semantic level in which the speaker's stance towards her locutory expression is manifested; similarly illocution belongs to a pragmatic level in which the speaker's stance towards her interlocutor is manifested, and finally attitude will be allocated to a socio-interactional conventionalized level. We believe that we must separate inferential clues that integrate the communicative activity – therefore, non linguistic factors, from the linguistic phenomena under study in order to achieve a more coherent description of the issues at hand.

2. Separating categories: an experimental rationale

In our view, modality belongs to a semantic level in which the speaker's stance towards her locutory expression is manifested; so the same illocution can be modalized differently, without affecting the illocutionary level. On the other hand, illocution belongs to a pragmatic level in which the speaker's stance towards her interlocutor is manifested; the illocution is the action the speaker is performing (order, question, assertion, calling, deixis, etc.). And finally, attitude is allocated to a socio-interactional conventionalized level in which the speaker shows her mood while performing a specific illocution (with a specific modality). The same illocution, that is, the same action, can be performed in different "ways", that meaning with different attitudes (seductive, irritated, tired, etc.). Paraphrasing Bally, we could say that attitude is the "Modus of Actum".

As stated above, we also believe that we must separate inferential clues that integrate the communicative activity – that is, non linguistics factors - from the linguistic phenomena under study, in order to achieve a more coherent description of the issues at hand. This means that the fact that a specific illocution can, in a specific context, be interpreted as a different action (let's say, a request interpreted as an order because of the hierarchic relationship between interactants or a question interpreted as a request because of a specific situation) does not depend on linguistic features. In this case the illocution performed is always a request or a question. This illocution plus inferential features may lead to a communicative interpretation as a different intention. We could explain this process in terms of Gricean implicatures (Grice 1975).

Cresti (2001) uses a commutation test to distinguish illocution and modality. We can extend this test and try to distinguish the three categories: attitude, illocution and modality. The main principle is that we cannot perform two things pertaining to the same categorical level at the same time. For example, we cannot perform a request and an order at the same time. If two things can be performed at the same time, it

means they pertain to different categories: so we can perform a request in a seductive way or in an irritated way, which means we can perform the same illocution with different attitudes. Also, we can perform the same request with different modalities. Following the same principle, we can perform different illocutions in an irritated way, but it seems we cannot show irritation and seduction at the same time, as these belong to the same category – i.e., attitude. Again, we can perform different illocutions with the same epistemic modality, for example, or the same illocution with different modalities, but we cannot perform a modality of certainty and a modality of uncertainty at the same time, or be epistemic and deontic at the same time.

Assuming the above mentioned definitions for the three different categories, what we aim at with the following experiment is to answer one main question: does prosody have the function to mark these different categories, and, if it does, how prosodic cues behave in order to mark modality, illocution and attitude?

Of course, the experiment is just the first step of a work in progress, and does not have any statistic value. Its ambition is just to show a possible research direction to study the relationship among these three categories and how they are linguistically marked.

2.1 Illocutions

As a start point, we can compare three different illocutions with the same locutive content. The locutive content is vem pro Brasil [(you) come to Brazil]. The three illocutions are: i) suggestion/recommendation; ii) invitation; iii) question. These three illocutions belong to the same class (directive) in Cresti´s repertory (Cresti 2000a; Cresti 2000b; Moneglia 2011), but it would be easy to extend the experiment with the same locutive content to illocutions of different classes (for instance to the assertive or expressive classes). Figure 1 shows the curve for the performance of suggestion/recommendation, figure 2 shows the curve for invitation, and figure 3 shows the curve for the question illocution. For each illocution the prosodic nucleus is circled. In order to make our argumentation clearer, we decided to circle two syllables, but it is probable that in most illocutions one syllable would be sufficient to identify the illocutive nucleus. To each figure, two sound files are associated: one features the whole utterance and the other features only the nucleus, thus allowing the verification that, in order for the specific illocution to be recognized, the nucleus is necessary and sufficient; the rest of the utterance represents just a preparation that has the function to host the locutive material beyond the few syllables that fulfill the nucleus. The fact that the nucleus is positioned in the last syllables (and in this case in the last tonic syllable) is not a necessary assumption for us. The position of the nucleus in the utterance depends on the type of illocution. Other illocutions may have the nucleus on the left or in the middle of the utterance. In illocutive terms, the

utterance can be built with three prosodic portions, just the nucleus being necessary: (preparation) – nucleus – (coda).

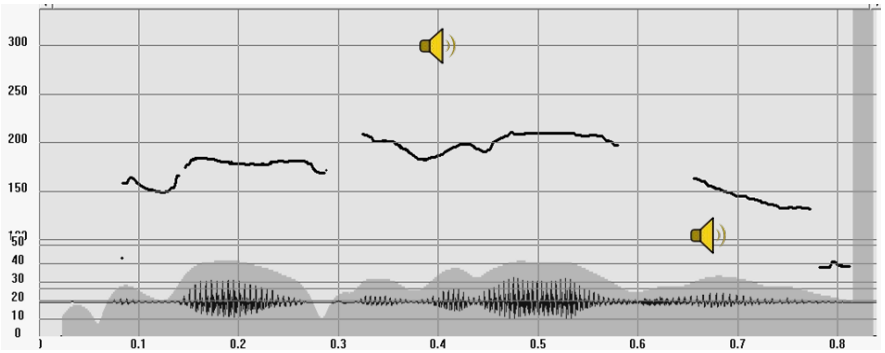


Figure 1. ‘Vem pro Brasil’. Illocution: suggestion/recommendation.

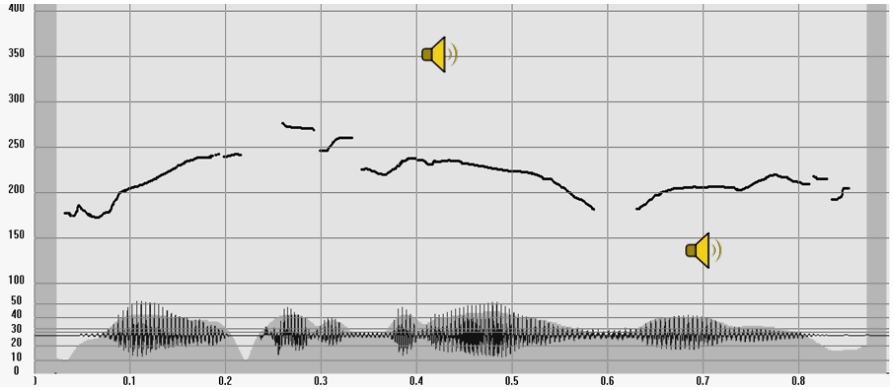


Figure 2. ‘Vem pro Brasil’. Illocution: invitation.

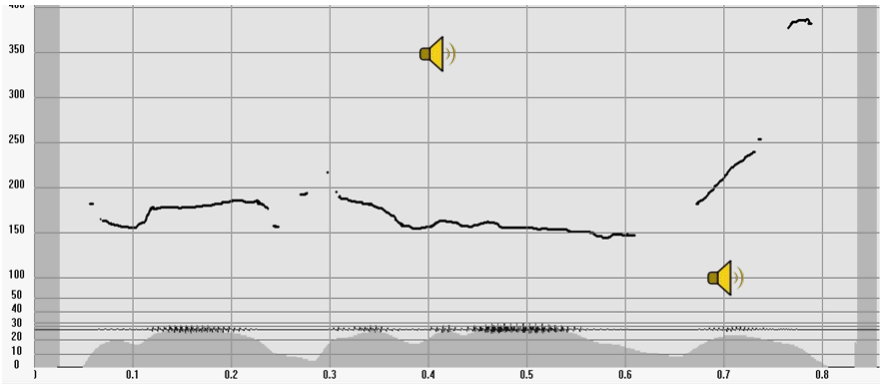


Figure 3. ‘Vem pro Brasil’. Illocution: question.

Looking at the three curves it is very easy to perceive the different form of the nucleus of the three illocutions, despite the fact that they belong to the same class. Someone could reasonably say that in the invitation illocution the first part of the utterance shows a sensibly different profile too, arguing that the illocutionary force cannot be attributed only to the circled part. But if we use the resynthesis and change the curve of the first part of the utterance, reassembling it to the first portion of the other two illocutions, we verify that in this part there is no functional movement with respect to illocution: in fact, the illocutionary value remains totally recoverable, as shown in figure 4 and its correspondent sound file. But we will come back to this later.

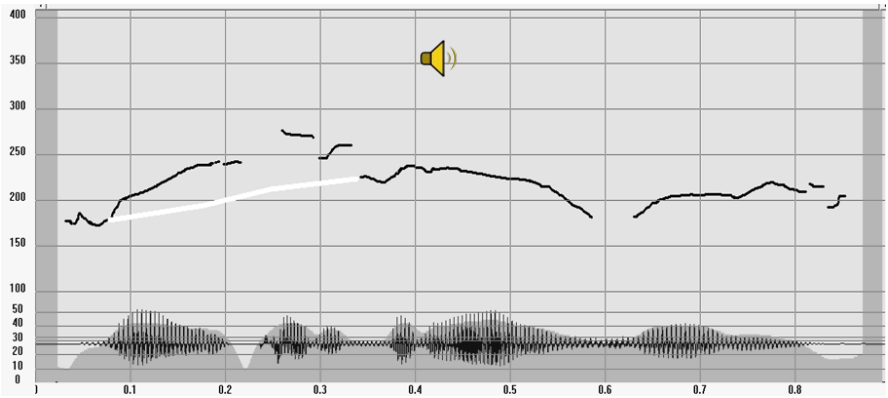


Figure 4. ‘Vem pro Brasil’. Illocution: invitation (resynthesis).

2.2 Modality

We can now take the same illocution of question and change its modality, in order to show that it does not give rise to any prosodic difference. Thus, in figures 5, the curve of the illocution of question with the modal verb ‘pode’ (*can*) is shown and in figure 6 the same illocution with the modal form ‘tem que’ (*must*) is shown. The two curves do not show any significant difference, and their prosodic realization can also be appreciated by listening to the audio files associated to them.

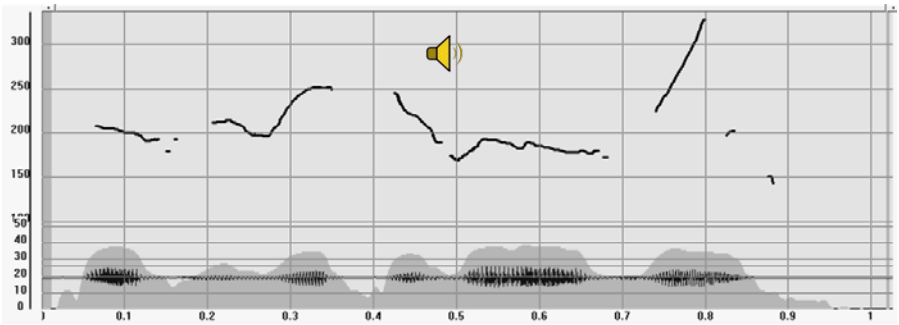


Figure 5. ‘Pode vir pro Brasil’. Illocution: question.

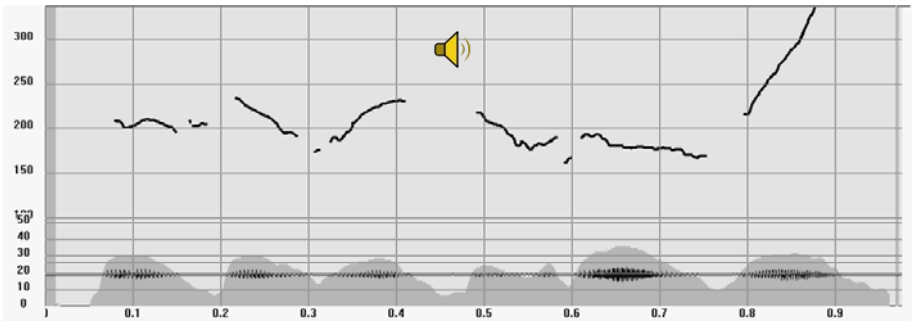


Figure 6. ‘Tem que vir pro Brasil’. Illocution: question.

Comparing the curves of figures 5 and 6 with that of figure 3, we can observe that the first part of the curve shows a little difference; the pode and the tem que forms feature a movement and a higher F0 level on the word vir if compared with the utterance without the modal lexeme of figure 3. In any case, it is remarkable that the curves with different modal verbs (poder and ter que) do not show any significant difference between themselves. So, the question is if the movement on the word vir, which in any case is not the modal lexeme, should be explained as a mark of modality or if it should be explained through different arguments. The major alternative arguments could be: i) the different syllabic dimensions, that induce a small variation of the F0 that does not have any functional value and is due to microprosodic factors (‘t Hart et al. 1990); ii) the different modal value, that induces a preferred realization with a change of attitude, which is not necessary and not functional with respect to modality in any case. In order to answer this question we employed resynthesis again and eliminated the movement on the word vir, as shown in figure 7.

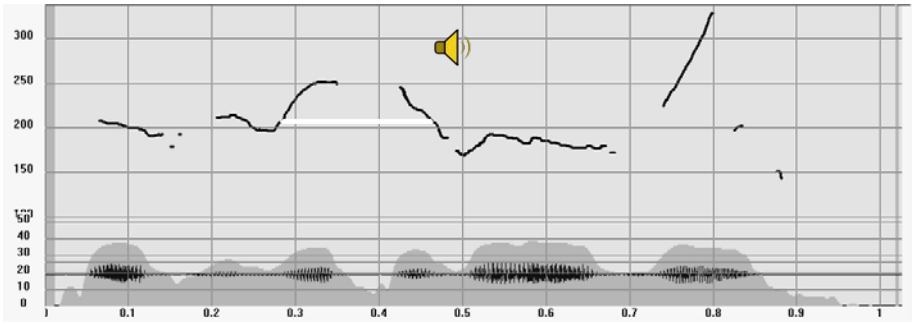


Figure 7. 'Pode vir pro Brasil'. Illocution: question (resynthesis).

Listening to the audio file associated to figure 7, we can verify that, even perceiving an acoustic difference with respect to the original utterance, the synthesized form is perfectly acceptable, and could be a possible realization of the same illocution with the same modality. This leads us to conclude that the different syllabic dimension of the utterance and maybe other factors induce a modification on the part of the curve that does not involve the illocutionary nucleus, but this modification does not depend directly on the modal value. The synthesized form does not lose the capacity to express the modal value of the original form at all.

2.3 Attitude

In order to show the influence of prosody toward attitude, in figures 8 and 9 we show the same illocution of question with an attitude that we could define as engaged and with an irritated attitude. We can thus compare three different attitudes for the same illocution expressed with the same locutive content, if we consider also the realization of figure 3, that we can define as normal or indifferent. In any case, the label we give to the three different attitudes is not important. What is important is to be able to recognize in them three different "ways" to perform the same illocution with the same locutive content.

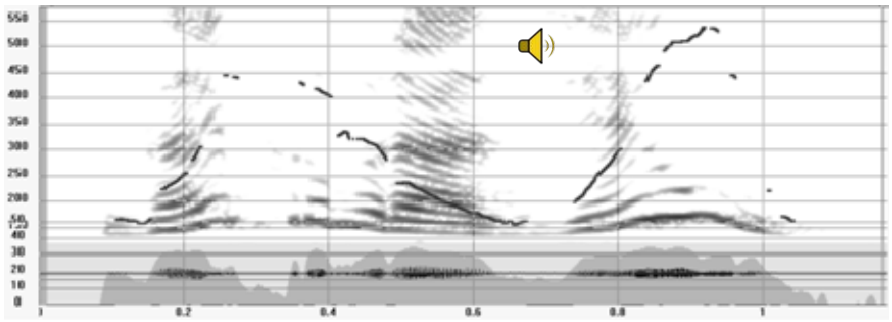


Figure 8. 'Vem pro Brasil'. Illocution: question. Attitude: engaged.

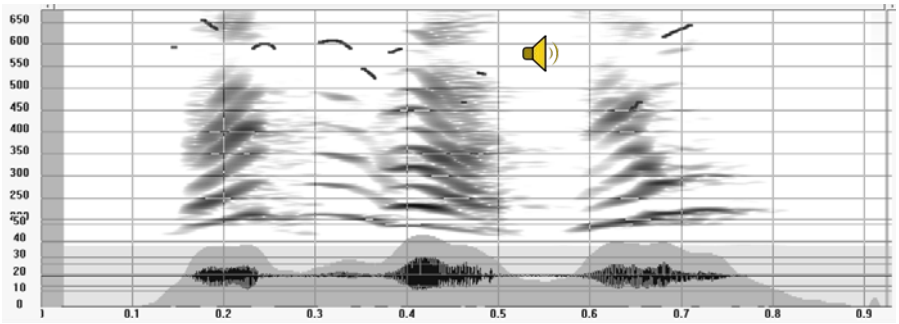


Figure 9. 'Vem pro Brasil'. Illocution: question. Attitude: irritated.

In order to observe the two new attitudes, we provide them in an alternate visualization format that allows appreciating the acoustic differences more closely. We can easily see that figures 3, 8 and 9 differ with respect to all relevant parameters (F0, intensity, duration), and that the differences cannot be confined to just a portion of the utterance. The whole utterance seems to be affected from the change of attitude. Taking the attitude of figure 3 as basis for comparison, we observe that the F0 in figure 8 is higher and shows a relevant movement also in the first portion of the utterance; the intensity looks higher, too. As far as the irritated attitude is concerned, both F0 and intensity are much higher than in the engaged attitude; also, we note that if there is no real difference in duration between the indifferent and the engaged attitudes, in the irritated attitude the duration is much shorter.

3. Attitude, illocution and modality: how they interrelate

3.1 The function of prosody

The experiment shown and commented in 2 leads us to some conclusions:

1. Prosody is an important cue in the rendition of illocution and attitude, but is not a constitutive cue of modality; if it is impossible to change illocution or attitude without modifying the prosodic parameters, in order to change modality we do not need to modify any prosodic parameter.
2. Given that it is a constitutive mark for both illocution and attitude, prosody affects the two categories in a very different way. While illocution is prosodically marked in a very short portion of the utterance, the nucleus, attitude is prosodically marked in the whole utterance. Illocution is recognizable by a specific form, that occupies one or very few syllables; on the contrary, the prosodic features of attitude are spread on the whole tonal unit, including the parts that constitute the preparation and the coda of the illocution.

In an attempt to better explain what we mean in reference to this last point, we can go back to figure 2, that shows the illocution of invitation, and to figure 4, that shows the same illocution modified by resynthesis, so as to have a more similar preparation to that of the other two illocutions (figures 1 and 3). The preparation, even after resynthesis, is not identical to that of the illocutions of suggestion/recommendation and that of question. In order to obtain the same result we should get an even flatter curve in the preparation. The result is shown in figure 10 and can be acoustically appreciated by listening to the associated audio file.

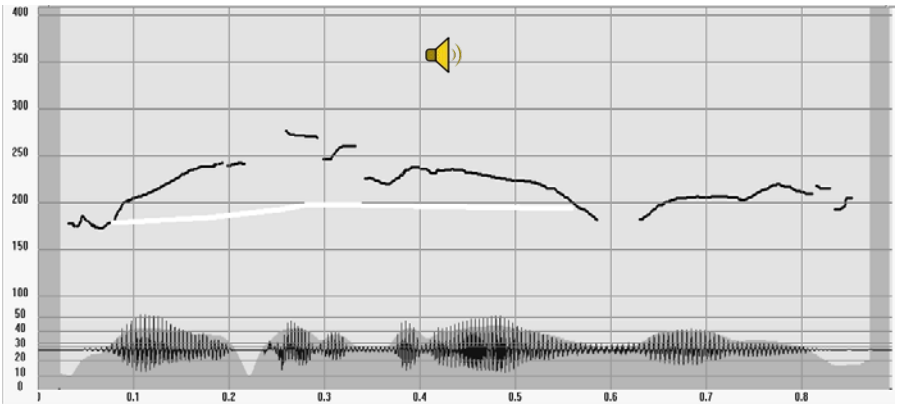


Figure 10. ‘Vem pro Brasil’. Illocution: invitation. Attitude: without enthusiasm.

The illocution is still clearly an invitation, but the attitude has changed. We can compare the original realization with the two resinthezised version and see how the attitude changes from an engaged invitation to less enthusiastic invitation and finally to an invitation without any enthusiasm whatsoever. This seems to demonstrate that:

1. without modifying the form of the nucleus we do not modify the illocution;
2. modifying a portion of the utterance that does not include the nucleus, we modify the attitude.

These conclusions, if confirmed, have important consequences for the study of the interface between prosody and the three categories that are the focus of this paper. The first consequence is that modality is not marked by prosodic cues; so, in order to study modality as defined here, we should not look for prosodic marks. The second conclusion is that prosody does mark both illocution and attitude, but these two categories are marked by prosody in very different ways: while illocution depends on the prosodic form of a small and easy to individualize part of the utterance, the prosodic cues that mark attitude are spread on the whole utterance. Again, if these conclusions can be confirmed, they show that it is perfectly possible to study the three categories without any confusion, but in order to do that, it is necessary to establish firstly a clear definition of what is under study, and secondly to look for the specific marks of each category. To measure the prosodic cues of the utterance without distinguishing between cues that are functional to illocutions and cues that are functional to attitude could only generate confusion and inconsistency.

3.2 The relationship among the three categories

But if we can distinguish the different relationship that these three categories have with prosody, we also should be able to understand how modality, illocution and attitude relate to each other. Some important questions are: what kind of relationship can we find among these three categories? Does a specific performance of one category determine or condition the others? If yes, in which direction and up to what point?

We believe that there is a sort of hierarchy among these three categories. Attitude is somehow superordinate to illocution and illocution is somehow superordinate to modality. But this does not mean that there is a deterministic relationship here. It means that it is more natural and probable that a certain attitude gives rise or associates itself to a certain type of illocution, and that one kind of illocution prefers a certain type of modalization. For example, it is easy to imagine an illocution of order with an irritated attitude, but the same attitude is less likely to be associated to an illocution of invitation, but, nevertheless, it is not impossible.

Similarly, we can easily expect an order to be modalized with must, while an invitation is more probably modalized differently, but it is also possible that it were modalized with must. Figure 11 represents our idea of the relationship among the three categories.

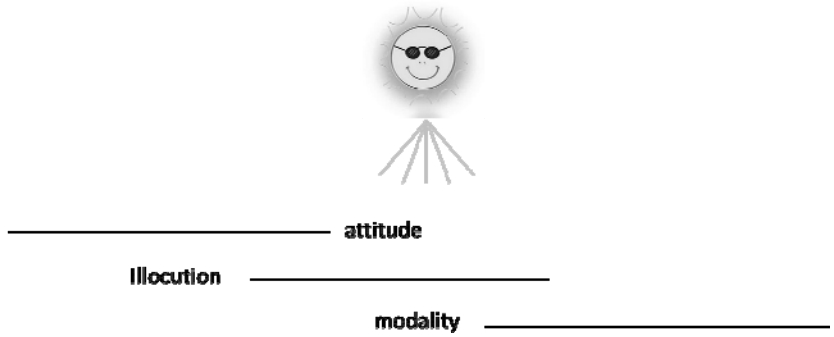


Figure 11. A possible scheme showing the relationship among attitude, illocution and modality.

What figure 11 intends to show is that the line of attitude projects a sort of shadow on part of the line of illocution, and that the line of illocution, similarly, projects a shadow on the modality line. This means that a light projected from above can reach any kind of attitude, but once the attitude is reached, only a light with a stronger and appropriate inclination can reach the shadowed portion of the illocution line, and, even more strongly, once the illocution is reached, only a very inclined light can reach the shadowed part of the modality line.

In respect to this, what happened in the experiment is interesting. We asked the same person to perform the three illocutions of suggestion/recommendation, question and invitation with the same locutive content. As we observed, the invitation illocution was performed with a specific attitude, different from that of question and suggestion. Probably this happened because an invitation, in order to be credible needs an engaged attitude. As it was shown in figure 10, the same illocution, with a preparation prosodically similar to the curve of question and suggestion, is still clearly an invitation, but with an attitude that sounds not enthusiastic at all, which is possible, but not normal for a credible invitation. What we want to point here is that the locutor, asked to perform a good invitation illocution, was naturally put in the condition to express an attitude appropriate to be credible. This should reinforce the argument that there is a relationship between attitude and illocution: some attitudes are more likely to be associated to an illocution, in our case invitation, but this is not necessary: in fact, the resynthesized form in figure 10 still features an interpretable invitation, but its attitude sounds very

different from that in figure 2, so that the invitation is interpreted as a non enthusiastic one.

Another aspect that we observed is that the same modal index can be interpreted with different modal values depending on the illocution in which it is placed. Figures 12 and 13 feature two different illocutions, respectively assertion and question, with the same locutive content: 'eu devo passar na casa dele'.

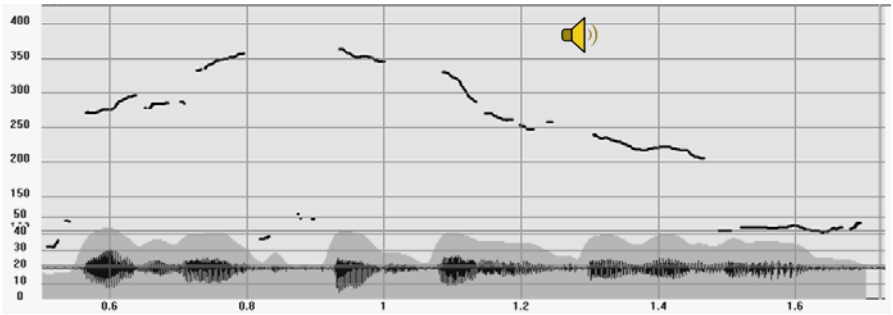


Figure 12. 'Eu devo passar na casa dele'. Illocution: assertion. Modal interpretation: epistemic.

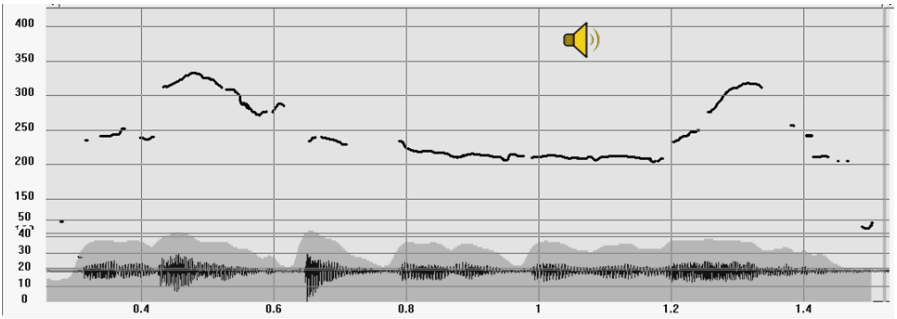


Figure 13. 'Eu devo passar na casa dele'. Illocution: question. Modal interpretation: deontic.

The modal value seems to receive two different interpretations depending on the illocution: in the assertion, the interpretation is 'I will probably stop by his place'; in the question, the interpretation seems to be 'do I have to stop by his place?' This test would confirm the idea that the interpretation of a modal index is driven, even if not determined, by the illocution performed.

4. Concluding remarks

As a conclusion, we propose that the study of the three categories of modality, illocution and attitude should, firstly, have consensus in the use of terminology and in a clear definition of the categories denoted by the three terms. Our proposal is that the term modality, defined with Bally as the "Modus of Dictum" should be considered a semantic category, the stance of the speaker toward the propositional content of his locution; on the other hand, the term illocution should apply to the action performed with an utterance; finally the term attitude should point to the way this action is performed, the "Modus of Actum".

We also propose that prosody does not play any role in marking modality, while it marks illocution and attitude, but in two very different ways: illocution is prosodically marked with the form of its nucleus, not affecting the rest of the locutive content of the tone unit; on the contrary, attitude is prosodically marked in the whole unit, but without changing the form of a specific illocution. Therefore we propose that there is a clear relationship among the three categories, as a specific attitude "prefers" some illocution and a specific illocution "prefers" some modalities, to the point that the same lexical index will receive a preferred interpretation due to the illocution it figures in.

There are other interesting questions linked to the discussion about modality, illocution and attitude that are outside the scope of this paper:

1. Up to what point is attitude conventionalized? We certainly can decide to have a seductive, or an irritated, or a lazy, or a tired attitude, but where is the border between these attitudes and non conventionalized emotions?
2. Up to what point the relationship among these three categories is a relation of probability (the superordinate category addresses the subordinate but does not determine it) and up to what point the superordinate category can bar the performance of a specific expression of the subordinate?
3. What is the relationship between the category of attitude and the illocutionary class of expressives? It seems that it is not easy to find the frontier between expressive illocutions (manifestation of surprise, irony, expression of wish, etc.) and attitudes. This should be better studied.
4. What is the scope of each category? We will not develop this point, but we believe that the scope of the illocution is the utterance, while the scope of modality and attitude is the information unit (Tucci 2006; Tucci 2009; Tucci 2010).

References

Austin, J.L. 1975. *How to Do Things with Words*. Oxford: Oxford University Press.

Bach, K. & Harnish, R.M. 1979. *Linguistic Communication and Speech Acts*. Cambridge, Mass.: MIT Press.

Bailly, G. & Holm B. 2002. Learning the hidden structure of speech: from communicative functions to prosody. In *Symposium on Prosody and Speech Processing*. University of Tokyo.

Bally, C. 1950. *Linguistique générale et linguistique française*. Berna: Francke Verlag. 1950.

Blakemore, D. 1992. *Understanding Utterances*. Oxford: Blackwell.

Bybee, J. 1985. *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: Benjamins.

Bybee, J. & Fleischman S. (eds) 1995. *Modality in Grammar and Discourse*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Cresti, E. 2000a. Per una nuova classificazione dell'illocuzione. In E. Burr (ed), *Atti del VI convegno SILFI – Tradizione e innovazione*. Duisburg: Cesati.

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

Cresti, E. 2001. Modalità e illocuzione. In P. Beccarla, C. Marello (eds), *Scritti in onore di Bice Mortara Garelli*. Torino: Edizioni dell'Orso.

Cruttenden, A. 1997. *Intonation* (2nd ed.). Cambridge: Cambridge University Press.

Green, M. 2009. Speech Acts. In E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2009 Edition). http://plato.stanford.edu/archives/spr2009/entries/speech-acts/

Grice, P. 1975. Logic and conversation. In P. Cole & J. Morgan (eds), *Syntax and semantics* vol 3. New York: Academic Press.

HUMAINE Emotion Annotation and Representation Language. http://emotion-research.net/projects/humaine/earl

Kärkkäinen, E. 1987. Towards a pragmatic description of modality. In I. Lindblad & M. Ljung (eds), *Proceedings from the Third Nordic Conference for English Studies*. Acta Universitatis Stockholmiensis LXXIII, 149-161.

Kiefer, F. 1994. Modality. In R.E. Asher (ed.), *The Encyclopedia of language and linguistics*. Oxford: Pergamon Press, 2515-2520.

Ladd, D. R. 1986. Intonational phrasing: The case for recursive prosodic structure. *Phonology Yearbook* 3, 311-340.

Leech, G. 1983. *Principles of Pragmatics*. London: Longman.

Local, J. 2005. Phonetic and interactional features of attitude in everyday conversation. Research proposal. http://www-users.york.ac.uk/~lang4/emotion-proposal.html

Lüdtke, H. 1980. *Kommunikationstheoretische Grundlagen des Sprachwandels*. Berlin: de Gruyter.

Mey, J. 1993. *Pragmatics: an Introduction*. Oxford: Blackwell.

Meyer, P.G. 1997. *Coming to know: studies in the lexical semantics and pragmatics of academic English*. Tübingen: Narr.

Mindt, D. 1998. *An empirical grammar of the English verb: modal verbs*. Berlin: Cornelson.

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

Moraes, J.A., Rilliard, A., Mota, B. & Shochi, T. Multimodal perception and production of attitudinal meaning in Brazilian Portuguese. In *Proceedings Speech Prosody 2010*. Chicago.

Mozziconacci, S.J.L. 2001. Emotion and attitude conveyed in speech by means of prosody. For the *2nd Workshop on Attitude, Personality and Emotions in User-Adapted Interaction*. Sonthofen, Germany.

Pierrehumbert, J. & Hirschberg J. 1990. The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan & M. Pollack (eds), *Intentions in Communication*. Cambridge, Mass: MIT Press, 271-311.

Ruthrof, H. 1991. Language and the Dominance of Modality. *Language and Style* 21: 315-326.

Schiffer, S. 1992. *Meaning*. Oxford: Clarendon Press.

Searle, J. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Schneider, S. 1999. *Il congiuntivo tra modality e subordinazione*. Roma: Carocci Editore.

Schubiger, M. 1958. *English Intonation: its Form and Function*. Tubingen: Niemeyer.

Shochi, T., Albergé V. & Rilliard A. 2006. How prosodic attitudes can be false friends: Japanese vs. French social affects. In *Speech Prosody 2006*. Dresden, Germany.

Sperber, D. & Wilson D. 1986. *Relevance: Communication and Cognition*. Oxford: Blackwell.

't Hart, J., Collier, R. & Cohen, A. 1990. *A Perceptual Study on Intonation: An Experimental Approach to Speech Melody*. Cambridge: Cambridge University Press.

Tucci, I. 2006. L'espressione lessicale nella modalità del parlato spontaneo: Analisi del corpus C-ORAL-ROM italiano. PhD diss., Università degli Studi di Firenze.

Tucci, I. 2009. The scope of lexical modality and the informational structure in spoken Italian. In L. Mereu (ed.), *Information structures and its interfaces*. Berlin/New York: Mouton, 203-226.

Tucci, I. 2010. La modalizzazione lessicale nell'enunciato. Semantica, pragmatica e composizionalità dei significati da una prospettiva corpus based. In *Atti del XLII Congresso SLI.* Pisa.

Ungeheuer, G. 1972. *Sprache und Kommunikation*. Hamburg: Helmut Buske Verlag.

FROM A PROSODIC POINT OF VIEW: REMARKS ON ATTITUDINAL MEANING

João Antônio de Moraes
Laboratório de Fonética Acústica, UFRJ/CNPq

To the memory of Ivan Fónagy,
with gratitude

1. Expressive intonation

There is a consensus that, along with a nuclear system of linguistic or "grammatical" intonation, languages make use of paralinguistic or "expressive" intonation (cf Ladd 1978; Ladd 2008; Gussenhoven 2004). Understanding how the expressive system interacts with the linguistic level is crucial to describing both the basic melodic patterns of a particular language and its expressive variants. However, to decide whether two melodic contours should be considered phonologically distinct or merely expressive variants of the same pattern is no simple task. It involves complex issues ranging from what acoustic parameters should be taken into account when describing the expressive prosody through to the definition of the various aspects of prosodic expressiveness itself. More precisely, are there distinct acoustic parameters to express grammatical and expressive prosody? In addition to the "classic" F0, intensity and duration, how important are parameters like voice quality and how relevant even are other channels (visual: gestures) for conveying expressive meaning? Moreover, would not phenomena usually seen to be expressive, like emotions, attitudes or feelings, display basically distinct prosodic behaviors, and be considered separately, as distinct types (or subtypes) of phenomena? Although highly relevant, these questions defy any clear-cut answer.

One (rather simplistic) way to decide whether two melodic contours should be seen as distinct patterns or as variants of a single pattern is to consider them as variants when the difference observed can be explained by the intervention of a gradient phenomenon, which does not disturb the overall configuration of the pattern in terms of a sequence of L and H tones. This view is taken explicitly, for example,

Mello H., Panunzi A., Raso T. (eds), *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation*

by Ward and Hirschberg (1988) when considering the melodic curves of the attitudes of uncertainty and incredulity variants of a L * + HLH% phonological pattern, although the scaling, that is, the level reached by the melodic H tone was clearly different in each case (figure 1).

The meaning distinction observed here, seen as a mere nuance (which is at best controversial), is then assigned to the expressive sphere. This approach is consistent with the view of the autosegmental-metrical (AM) theory, for which:

> the tonal span of melodic accents expands or contracts, in the phonetic component, according to the speaker's involvement, in such a way that the more emphatic a statement, the more the range of tonal inflection increases (Prieto 2003: 28).

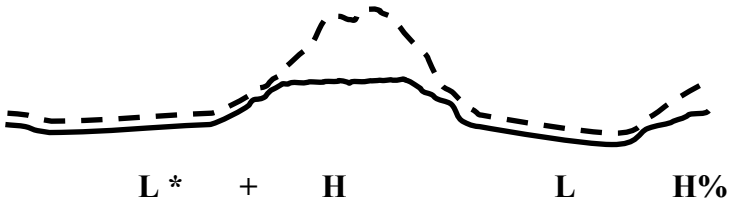


Figure 1. Superposition of the uncertainty (solid line) and incredulity (dashed line) pitch contours, produced with the same sentence 'Eleven in the morning', analyzed as the same phonological contour L * + HLH%, based on Ward and Hirschberg (1988).

It is assumed that pitch range (the magnitude of F0 excursions and/or register) is a gradient phenomenon, and thus essentially expressive, and should therefore not be represented as phonological, since it belongs only to the «phonetic component, and does not substantially affect the linguistic meaning» (Prieto 2003: 20)[1].

It turns out that the so-called expressive phenomena are not restricted to cases of emphasis, in which there is in fact greater speaker involvement with what is being said. As a matter of fact, correlating "expressive patterns" to "gradient" and "grammatical patterns" to "discrete" does not always work.

To approach this issue better, it is worthwhile to look first at the typology of affective phenomena.

[1] There are, however, proposals to integrate some kind of gradient variations to the AM theory (Ladd 1983), and even attempts to represent emotional speech using ToBI notation (Stibbard 2000).

2. Types of affective states

In fact, the general concept of prosodic expressivity covers the manifestation of various categories of affective state. Léon (1993) proposes a continuum of five steps, from raw emotion, to emotion-feeling, intellectual emotion (or attitude), linguistically encoded emotion and, finally (now outside the expressive domain), grammatical modality, which could be represented graphically as:

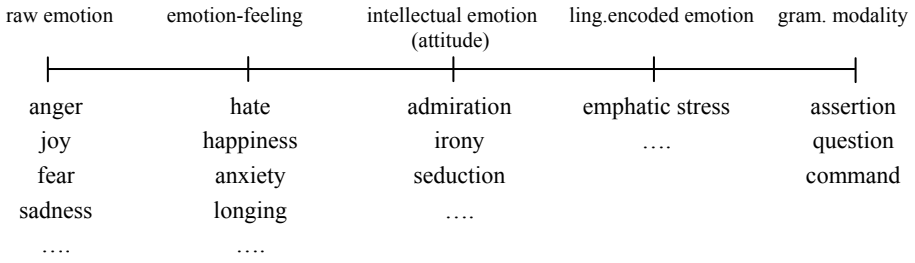


Figure 2.

Scherer (2000) and Scherer & Bänziger (2004) have proposed a detailed design feature approach to distinguish five classes of affective state: emotions (e.g., angry, sad, joyful, fearful, ashamed, proud, elated, desperate), moods (e.g., cheerful, gloomy, irritable, listless, depressed, buoyant), interpersonal stances (e.g., distant, cold, warm, supportive, contemptuous), preferences/attitudes (e.g., liking, loving, hating, valuing, desiring) and affect dispositions (e.g., nervous, anxious, reckless, morose, hostile). This typology is based on the behavior of seven parameters, rated in three degrees, H(igh), M(edium) and L(ow), as can be seen in the table below:

Table 1.

Types of affect / Design features	emotions	moods	interpersonal stances	preferences/ attitudes	affect dispositions
intensity	H	M	M	M	L
duration	L	M	M	H	H
synchronization	H	L	L	L	L
event focus	H	L	M	L	L
appraisal elicitation	H	L	L	L	L
rapidity of change	H	M	H	L	L
behavior impact	H	L	M	M	M

From a strictly prosodic perspective the main concern is to establish the extent to which these categories show different prosodic behaviors, that is, whether there are prosodic features which characterize these different categories of affective states,

either because the categories preferentially use different parameters, or the same parameters in different ways, locally or globally. Another unresolved issue is to determine to what extent they can combine or whether, on the contrary, they should be seen as mutually exclusive, and belonging to the same paradigm.

3. The emotion vs attitude distinction

In the tradition of intonational studies, only two types of expressive phenomena are usually distinguished: the vocal expressions of emotions, on the one hand, and those of the speaker's attitudes, on the other (Fónagy 1993; Fónagy 2000; Fónagy 2006; Couper-Kuhlen 1986).

Couper-Kuhlen (1986: 185-7) argues that emotion is an "inner state" or "feeling" of the speaker (for instance, the speaker is bored, impatient, anxious, happy), while attitude is a kind of behavior toward the interlocutor (for instance the speaker is being friendly, arrogant, sexy, polite). She draws attention to the fact that the first group can be paraphrased by: «X (the speaker) is [for instance, happy] (in uttering p)», and in the second: «X is being [for instance, polite] (in uttering p)». It is worth noting that here the speaker's "inner state" (or feeling) covers three of Scherer's categories of affective states, namely emotion (sad), mood (bored), and affect disposition (anxious), while the speaker's behavior corresponds to Scherer's interpersonal stance.

Emotions, or at least basic, primary ones like anger, joy, fear and sadness are seen as «spontaneous discharges of psychic tension» (Fónagy 1993: 27). Their manifestations are largely universal and they correlate with physiological changes that affect the vocal tract as a whole.

According to Fónagy, there are two levels of symbolism involved in emotions: direct laryngeal gestures (voice quality) and indirect tonal gestures (melody) (Fónagy & Bérard 2006: 22-24). Attitudes, on the other hand, are conventional and act mainly at the glottal level (Fónagy 2000: 138).

Intuitively we feel that an emotional state, working as an independent, orthogonal system, can be added to any language production, whatever the speech act[2]. Accordingly, a declarative or interrogative sentence can be uttered in several emotional states, as in figure 3.

[2] Although some marginal conflicts or restrictions between certain combinations of emotion type and kind of speech act have been observed. (Colamarco & Moraes 2008; Colamarco 2009).

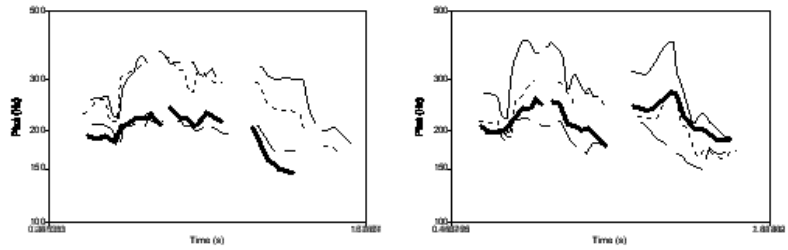


Figure 3. Superposition of the pitch contours of the sentence 'Roberta já sabe' [Roberta already knows] uttered with three different emotions, as an assertion (left panel) and as a yes-no question (right panel): joy (solid line), anger (dotted line), sadness (broken line), and neutral contour (thick line); female speaker (duration normalized). From Colamarco (2009).

The main change observed here affects the utterance as a whole (register and pitch span), hence the difficulty or even impossibility of representing the observed differences in a ToBI-like system. In addition to the F0, the parameter represented in figure 3, emotions are also expressed through other vocal parameters such as intensity, duration and voice quality, and even through other (e.g., visual) channels (gestures, especially facial ones).

4. Ambiguity of the term attitude

In contrast with emotions, attitudes correspond to «a controlled behavior, with a moral and intellectual component» (Fónagy 1993); rather like "socially tamed" emotions. Typical attitudinal labels are complaint, irony, politeness, longing.

However, defining and delimiting the field of attitude precisely is a difficult task, as the term is particularly ambiguous. Indeed, two basic uses of "attitude" that are of direct concern to the study of prosody are the speaker's attitude towards his interlocutor (the so-called social or interpersonal attitude) and the speaker's attitude to what is being said (the propositional attitude) (Moraes et al. 2010)[3,4].

[3] Wichman (2000), for instance, also distinguishes, from the intonational perspective, propositional attitudes («attitudes towards propositions (...) which are functions of opinions, knowledge or beliefs» [e.g. impressed, disapproving ...]) from attitudes *tout court*, which coincides with attitudes that we are calling social or interpersonal (attitudes as «speaker behaviour in a given situation, either as intended by the speaker, or as inferred by the receiver, or both») [e.g. condescending, rude...]). Contrary to what is done here, however, she classifies the first type along with the emotions, and considers them as being expressed by what she calls "expressive" intonation (in a quite personal use of this term); these attitudes would be centred in the speaker and do not depend crucially on the interaction. The second type, referred to as attitude *tout court*, is manifested, according to the author, by so-called attitudinal intonation, and is directly related to the presence of the interlocutor.

4.1 Social or interpersonal attitudes

Social or interpersonal attitudes are speaker's attitudes towards his interlocutor[5]. Typical attitudes of this kind are seduction, hostility, politeness.

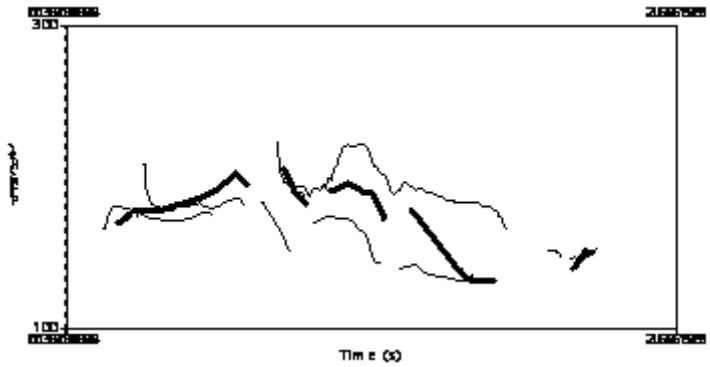


Figure 4. Superposition of the pitch contours of the sentence 'Roberta jogava' [Roberta used to play/Roberta was playing] uttered with a seductive (top line), indifferent (bottom line) and neutral (middle thick line) attitudes; female speaker (duration normalized).

Studies that investigate the manifestation of social attitudes in language often include attitudes such as: friendly/aggressive, patient/impatient, authoritarian/ submissive, seductive/indifferent, polite/impolite, or even sure/unsure, shy/outgoing, tense/calm, which on Scherer's criteria would be classified rather as affective dispositions.

Social attitudes display general prosodic behavior somewhat comparable to that of emotions, in the sense that, in both, the melodic changes over the basic pattern tend modify the utterance as a whole (register/span) (figure 4), although vocal quality modifications clearly play a less important part in social attitudes than in emotions.

---

[4] It is worth noting that sometimes a single label, like irritation, can apply to both sets of categories: one can be irritated towards an interlocutor: É a terceira vez que eu te dou essa informação!!! [It's the third time I've given you this information!!!], or irritated by the very fact expressed in the propositional content of the utterance: A conta deu errado de novo!!! [The sum has gone wrong again!!!].

[5] In the tradition of social psychology (Osgood & Tzeng 1990) the expression "social attitude" acquires a broader sense: a negative, positive or neutral stand for a given "object" (person or event).

4.2 Propositional Attitudes

In the tradition of semantics and philosophy of language (Russell 1918; Quine 1956), a propositional attitude denotes a mental state (posture) relating the speaker to a proposition (not to another person or event). Examples are belief (in its truth), desire (that what is expressed in the proposition will occur), hope etc. Therefore, a given proposition (or *dictum*, in Bally terms) can be uttered in different ways (*modus*) in a cognitive, volitional or emotional light.

ATTITUDINAL VERBS		PROPOSITION
I believe, doubt, deny, accept, know, think desire, expect, hope…	that	[tomorrow will rain]

As Aydede (2010) puts it:

> Propositional attitudes are the thoughts described by such sentence forms as 'S believes that P', 'S hopes that P', 'S desires that P', etc., where 'S' refers to the subject of the attitude, 'P' is any sentence, and 'that P' refers to the proposition that is the object of the attitude. If we let 'A' stand for such attitude verbs as 'believe', 'desire', 'hope', 'intend', 'think', etc., then the propositional attitude statements all have the form: S As that P.

Of course, when the propositional attitude is expressed intonationally, only the original proposition (P) of that formula (S As that P) remains. That is, the subject (always in the first person, in this case), the propositional attitude verb and the that particle (the S As that part) are "replaced" by the melodic contour. They are no longer reported attitudes (McKay & Michael 2010), but necessarily genuine speaker's attitudes.

From a prosodic point of view, the crucial point is to know how many (and which) dedicated, propositional-attitudinal melodic patterns there are. This is a controversial matter, but there is no doubt that some typical propositional attitudes can be (and are) expressed through prosody. Examples of such propositional attitudes are irony, uncertainty, and incredulity.

Another interesting point is that, to some extent, the concept of propositional attitude is very close to, and even overlaps with, that of speech act (Searle 1969; Vanderveken 1990), or at least with that of illocution (Cresti 2000; Firenzuoli 2003). For instance, "deny" or "accept" are considered to be propositional verbs as well as performative verbs characterizing specific speech acts.

4.2.1 *Propositional attitudes and speech acts*

According to Searle's Speech Acts (SA) Theory (Searle 1969; Searle & Vanderveken 1985), the attitudes of belief, desire and intention have a special status, since they typically correspond to mental states, which in turn establishes the so-called *sincerity conditions* of assertive, directive and commissive acts, respectively. Thus the attitude/mental state of belief in the truth of the propositional content must implicitly be present in an assertion; the attitude of desire that the act be performed must be present in a request; and the intention of performing the action being promised must be present in a promise.

If it seems to make little sense to talk in general terms about an intonation of belief or of desire, it is perhaps easier to accept the existence of melodic contours relating to different degrees of belief, desire or commitment.

In fact the SA theory proposes a componential analysis of speech acts; the *illocutionary force* of a speech act is not considered a primitive notion, but it depends on six dimensions or components, namely, (i) an illocutionary point, (ii) the mode of achievement of the illocutionary point, (iii) the propositional content, (iv) preparatory and (v) sincerity conditions and, finally, (vi) the degree of strength (Vanderveken 1990: 103). At the interface between propositional attitudes and intonation, the notion of strength is especially relevant, as noted by Reis (2007). Thus the mental states which determine the sincerity conditions of speech acts can be expressed with different degrees of strength, depending on the illocutionary force; so in supplication there is a stronger attitude of desire than in a request, and belief is stronger in a testimony than in a conjecture (Vanderveken 1990: 119).

Note that, unlike what happens in assertive, directive and commissive acts, propositional attitudes are intrinsically constitutive of *expressive speech acts*, whose illocutionary point «consists of expressing propositional attitudes of the speaker about a state of affairs» (Vanderveken 1990: 105). In such acts there is no "neutral" condition of sincerity, as in non-expressive acts, in the sense that every expressive illocutionary force necessarily has a special sincerity condition, represented in each case through the use of adjectives like 'I am glad that…' (or 'How glad I am that ...'). Typical propositional attitudes relating to expressive speech acts are approval, disapproval, sorrow, joy, sadness, sympathy, gratitude, and regret.

4.2.2 *Belief attitude in statements and questions*

A propositional attitude such as belief (in the truth of a given propositional content) is indeed quite productive in intonational terms and can surface, not just in a binary opposition (certainty *vs.* doubt), but rather in differing degrees, that may be arranged in a *continuum* (Moraes 2008b; Reis 2010). Thus, in so-called assertive sentences, the different melodic propositional attitudinal contours may be displayed on an axis representing the speaker's degree of certainty / uncertainty (commitment to the truth) towards the veracity of the expressed propositional content (PC). On a

continuum ranging from certainty that a given propositional content is true to certainty that it is false, through the neutral point, which would be doubt, we could point to at least five melodic patterns that are typical of Brazilian Portuguese (BP). These range from corrective emphasis, in which the speaker strongly asserts the truth of P, to irony which, on the contrary, denies the truth of P, through obviousness, neutral assertion, and incredulity, as shown in the diagram below (for an explanation of my use of these labels, see Moraes 2008a):

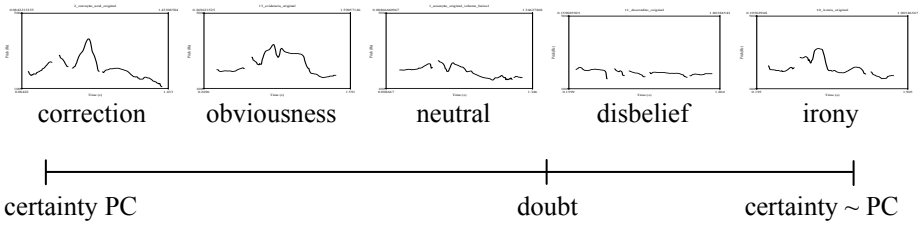


Figure 5.

Note that (i) I included denial of the certainty of PC (irony) on the same semantic axis, with doubt occupying, not the extreme, but an intermediate position; and (ii) the melodic form varies in a discrete, rather than a gradient, manner (Moraes 2008a), which underlines the conventional, language-specific nature of propositional attitudinal intonation patterns (Moraes et al. 2010). Even a superficial visual examination reveals, for instance, that disbelief and irony show very distinct melodic behaviors, despite their semantic proximity, and that irony and correction are more similar to each other than to the neutral pattern.

Similarly, attitudinal melodic patterns usually classified as belonging to the class of yes-no interrogatives can be analyzed as containing different "concentrations" of the attitudes of certainty and doubt (knowledge/expectation of the answer). Thus, at one end of the *continuum* we have the confirmation-seeking yes-no question, marked by the expectation of a response that confirms the PC (a positive polarity: the speaker assumes the truth of P with a reasonable degree of certainty) and at the other end, the rhetorical yes-no question, which assumes precisely the opposite[6], through the neutral question, where there is no clear polarity

[6] In BP typical rhetoric questions display reverse polarity, that is, if there is a negation in the PC, the implication is positive, if there isn't such a negation, the implication is negative. This is true for both, yes-no and wh-rethorical questions: Eu já não te expliquei isso? [Didn't I explain this to you before?] (implying → Yes, I did); Você gosta de levar bronca? [Do you like to be scolded?] (implying → No, you don't). Quem não gosta de elogio? [Who does not like compliments?] (implying → Everyone does). Quem gosta de ser repreendido? [Who

(there is typically doubt) and the incredulous question, in which the speaker assigns little probability of the PC being true.

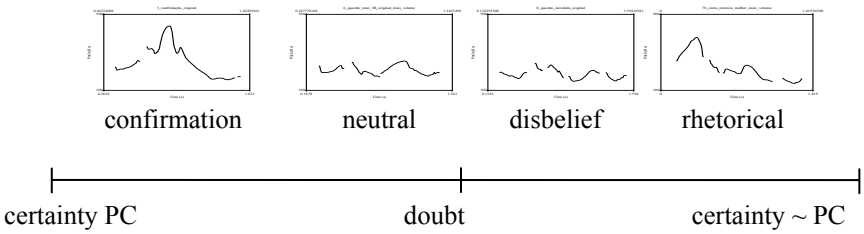


Figure 6.

If we agree that in the "typical" assertion there is necessarily a commitment on the part of the speaker to the truth of the propositional content being expressed (which, in Searle's view, corresponds to the *sincerity condition*), that commitment obviously ceases in, for example, an ironic assertion or even in a statement expressing doubt. Similarly, if in the typical question the *preparatory condition* for its success is the speaker's not knowing the answer, in a confirming or rhetorical question, or an incredulous question, this condition is violated, originating another speech act.

Unlike what happens with emotions and social attitudes, the pitch contours associated with propositional attitudes clearly show more substantial and punctual, localized melodic configurations (figure 7), and contrasting melodic patterns (Moraes & Rilliard in preparation).

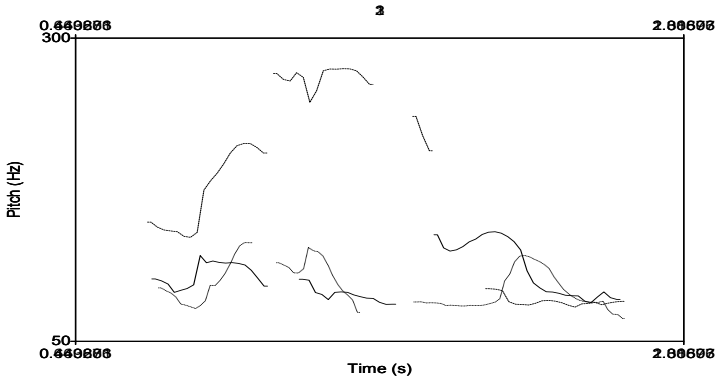


Figure 7. Superposition of the pitch contours of the yes-no question 'Roberta dançava?' [Did Roberta use to dance? / Was Roberta dancing?] uttered with a neutral attitude (solid line), a confirmative attitude (broken line) and an incredulous attitude (dotted line); male speaker.

---

likes to be scolded?] (implying → No-one does). In BP, only yes-no questions show different melodic patterns when employed as rhetoric or as "real" questions.

5. Social *vs.* propositional attitudinal prosody: perception and production

Recent studies (Moraes et al. 2010; Moraes et al. 2011 submitted; Moraes et al. in preparation) as part of the PADE Project[7], have shown that, in BP, propositional and social attitudes in fact display differentiated prosodic behavior in both perception and production.

5.1 Perception

Thus, Moraes et al. (2010) examines production and perception involving six social attitudes (arrogance, authority, seduction, contempt, irritation and politeness), and five propositional attitudes (doubt, obviousness, disbelief, irony and surprise), all expressed through the neutral declarative sentence 'Roberta dançava' [Roberta was dancing/Roberta used to dance].

In Moraes et al. (2011 submitted), the same sentence, uttered as a yes-no question 'Roberta dançava?' [Was Roberta dancing?/ Did Roberta use to dance?], was spoken with the same six social attitudes and with four propositional attitudes, namely, confirmation, strangeness, rhetoricity and surprise. Both studies also included the so-called "neutral" (respectively, assertive or interrogative) attitude.

Two Brazilian speakers were recorded and filmed while producing these sentences. The resulting audio and visual stimuli were submitted to an identification (forced choice) test with 30 subjects, who had to identify the speaker's attitude from the audio alone, the image alone and, finally, from both information sources simultaneously.

The order in which the stimuli were presented was balanced: half the subjects judged video stimuli first and then audio stimuli (and finally both together), while the other half did things the other way round. Subjects listened to/viewed the stimuli and gave their answers on a computer screen using a slider which, in addition to indicating the attitude chosen, also reported the relative intensity of the perceived attitude on a scale from 0 to 100.

The results for both modalities show not only that the propositional attitudes were in general significantly better recognized than social ones, but more

[7] The goals of the PADE Project, under the direction of Albert Rilliard (Rilliard 2010), include examining attitudinal prosody cross-linguistically in languages such as French, Japanese, American English and Brazilian Portuguese, assessing the specific weight of visual and audio channels in its manifestation (Shochi et al. 2007; Rilliard et al. 2009; Moraes et al. 2010)

specifically that the visual channel plays a much more important role than audio in recognition of social attitudes (Figures 8 and 9).

Specifically for assertions, the audio channel for propositional attitudes returned a score of 61% correct answers (much higher than the 14% chance level), while for social attitudes it produced average recognition of only 25% (close to the 17% chance level for this case). Although the contribution of the visual channel is very important in both, it is crucial in relation to social attitudes, which are indeed visually dependent.

In interrogatives, almost the same results were obtained for audio stimuli: 60% for propositional and 28% for social attitudes, with the visual channel contributing less in relation to the propositional attitudes.

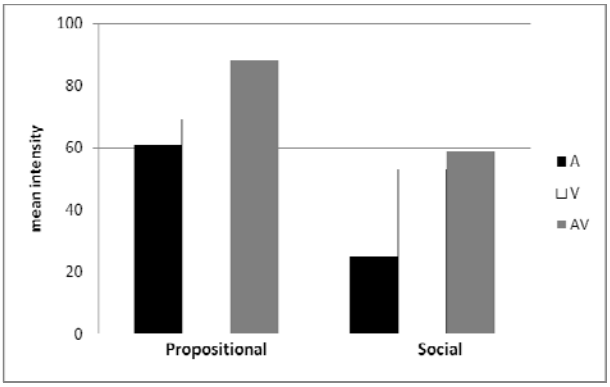


Figure 8. Assertive sentences: mean intensity of correct answers in each condition, for propositional and social attitudes, both speakers. A stands for audio condition, V for video and AV for both together.

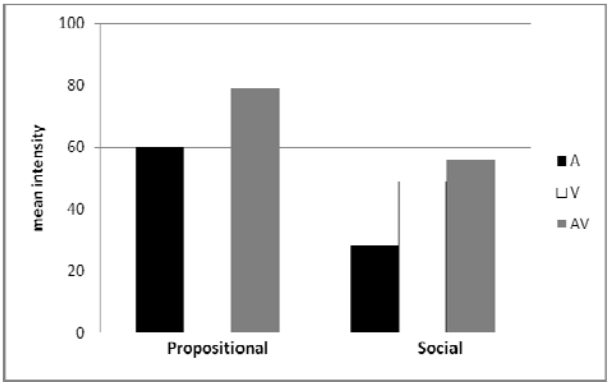


Figure 9. Interrogative sentences: mean intensity of correct answers in each condition, for propositional and social attitudes, both speakers. A stands for audio condition, V for video and AV for both together.

5.2 Production

The assertive sentence with neutral attitude can be characterized melodically by a moderate F0 fall in the final, nuclear position, specifically between the last pre-stressed and stressed syllables, which also assumes a falling internal configuration.

Looking at how social attitudes surface in melodic terms, one sees that they show rather subtle melodic distinctions (figure 10), and that the neutral contour is basically preserved. Figure 11 shows the same F0 contours as in figure 10 after stylization to eliminate perceptually irrelevant melodic modulations (‘t Hart et al. 1990), which makes the great similarity between the patterns even more evident.

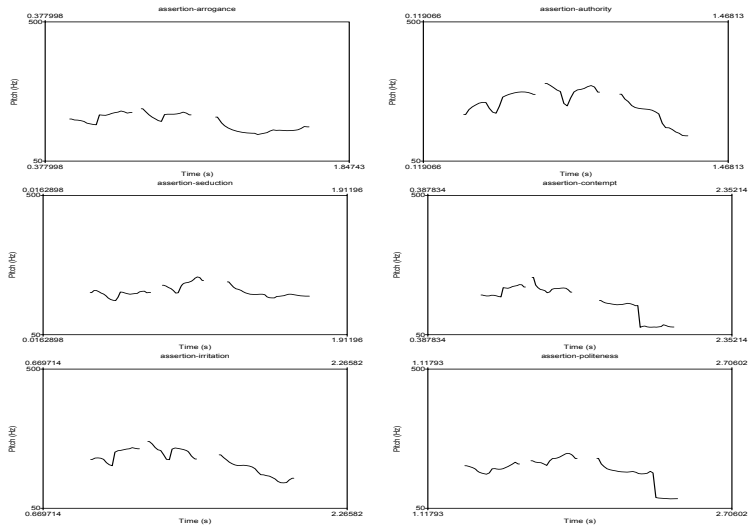


Figure 10. Pitch contours of the assertive sentence ‘Roberta dançava’ [Roberta was dancing/ Roberta used to dance] uttered with six social attitudes, female speaker. From top to bottom: arrogance and authority; seduction and contempt; irritation and politeness.

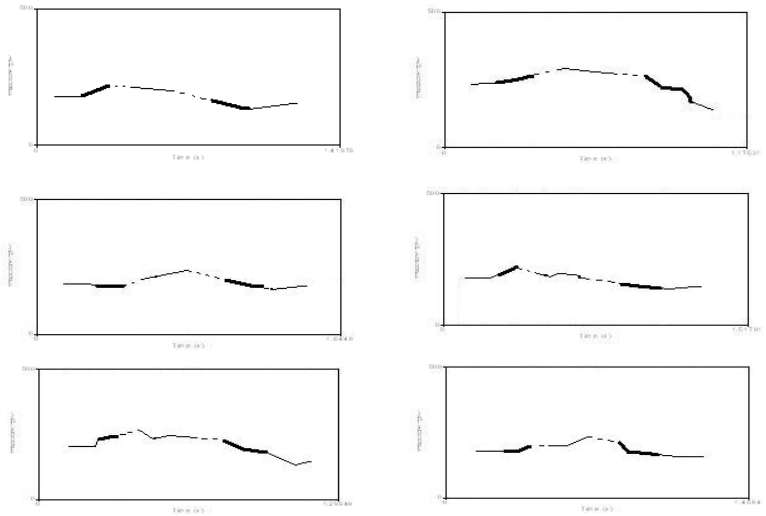


Figure 11. Stylized pitch contours of the assertive sentence 'Roberta dançava' [Roberta was dancing/ Roberta used to dance] uttered with six social attitudes, female speaker. The thicker line indicates the stressed vowels, the dotted line, voiceless consonants. From top to bottom: arrogance and authority; seduction and contempt; irritation and politeness.

On the other hand, most of the propositional attitudes examined here (figures 12 and 13) show important, punctual changes in the melodic contour, which modify its basic configuration; that is why they are better perceived by the ear. These changes are located mainly in the nuclear position, more specifically the last stressed syllable, and/or in the contrast between this syllable and the preceding one. The tonal importance of the nuclear position has been confirmed by manipulating the F0 at specific points in the melodic patterns of propositional attitudinal utterances, then validating by perception tests (Moraes 2008a).

Accordingly, in disbelief, both nuclear syllables are produced at a very low melodic level; in obviousness, the last stressed syllable is produced at quite a high level (for an assertive sentence); in irony the last stressed syllable assumes a typical, circumflex (rising-falling) shape; and doubt displays – among other things – a high last pre-stressed syllable. In addition in the duration level, irony, disbelief and doubt also display greater duration in general, especially a lengthening of the last stressed syllable. These major differences between the expression of social and propositional attitudes are observed among interrogatives as well.

The propositional (assertive or interrogative) contours can be represented more easily by an AM notation system, such as ToBI.

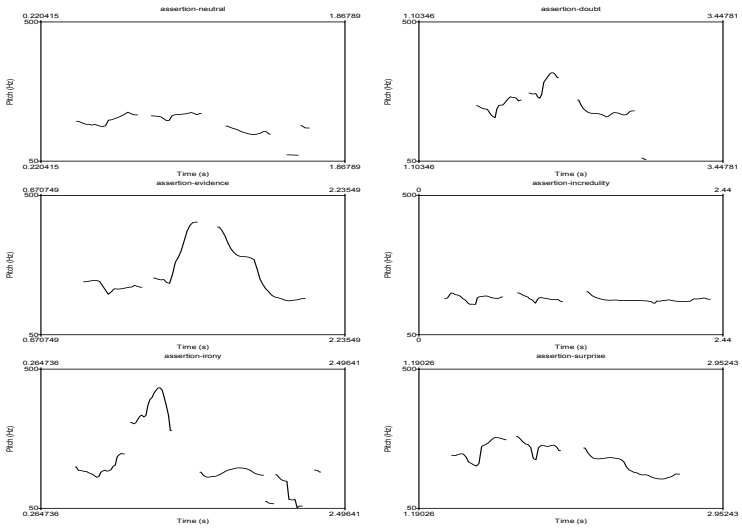


Figure 12. Pitch contours of the assertive sentence 'Roberta dançava' [Roberta was dancing/ Roberta used to dance] uttered with neutral and five propositional attitudes, female speaker. From top to bottom: neutral and doubt; obviousness and disbelief, irony and surprise.

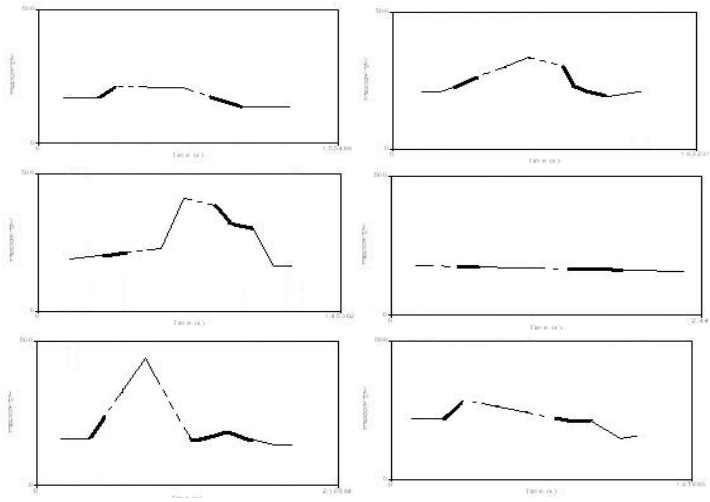


Figure 13. Stylized pitch contours of the assertive sentence 'Roberta dançava' [Roberta was dancing/ Roberta used to dance] uttered with neutral and five propositional attitudes, female speaker. The thicker line indicates the stressed vowels, the dotted line, voiceless consonants. From top to bottom: neutral and doubt; obviousness and disbelief, irony and surprise.

6. To conclude

The results of perceptive analysis (Moraes et al. 2010 submitted), acoustic analysis (Moraes et al. in preparation) and even F0 manipulation experiments with resynthesis (Moraes 2008a) reinforce the idea that there are two independent prosodic systems: emotions + social attitudes vs. propositional attitudes (+ speech acts).

In the original scheme proposed by Aubergé (2002), the attitudinal functions are located halfway between the linguistic and non-linguistic functions. The proposal here is then to split the two categories of attitudes, putting social attitudes together with emotions, and propositional ones with speech acts (in italics in the scheme).

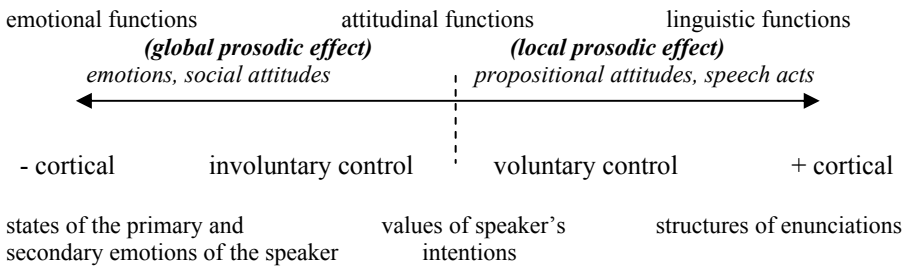


Figure 14.

Emotions and social attitudes do not conflict with speech acts or propositional attitudes: in fact they can be added to them without destroying the basic communicative value. Also, from a prosodic perspective, neither do they significantly disturb the basic melodic pattern - in fact, the pattern is largely preserved; to be more precise, it becomes a variant of the original (unmarked) pattern.

This means that the phonological representation of a particular illocutionary act spoken with different emotional or social-attitudinal values would be the same: there are no localized, punctual F0 changes, but global modifications in the overall pattern (register and tonal span), not to be represented in phonological form. With propositional attitudes and speech acts, the changes are local (discrete), leading to distinct phonological analyses.

Finally, regarding the participation of different "media" in the expression of affective meaning, our data reveal that the visual channel (facial stimuli) contributes more to the production and perception of social attitudes than the audio channel (prosody and voice quality). The same holds for emotions, as Levitt (1964) has shown in his classic study. For propositional attitudes, the opposite occurs,

confirming the view of Pakosz (1983: 321): "speakers tend to rely more heavily, in the expression of some affects, on one channel".

The table below summarizes our view of the involvement of different parameters in the expression of emotions, social and propositional attitudes. This is to be tested in a future study, particularly with regard to voice quality[8].

Table 2.

Channel / Category	Gestures	Voice Quality	Prosody
Emotion	++++	+++	+
Social Attitude	++	+	+
Propositional Attitude	+	-	+++

References

Aydede, M. 2010. The Language of Thought Hypothesis. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2010 Edition). http://plato.stanford.edu/archives/fall2010/entries/language-thought

Aubergé, V. 2002. A gestalt morphology of prosody directed by functions: the example of a step by step model developed at ICP. In *Proceedings of Speech Prosody 2002*. Aix-en-Provence.

Colamarco, M. 2009. A expressão das emoções em atos de fala no português do Brasil: produção e percepção. Master diss., Faculdade de Letras, UFRJ.

Colamarco, M. & Moraes, J. 2008. Emotion expression in speech acts in Braziliam Portuguese: production and perception. In *Proceedings of Speech Prosody 2008*. Campinas, Brazil, 717-720.

Couper-Kuhlen, E. 1986. *An Introduction to English Prosody*. London and Tuebingen: Edward Arnold and Niemeyer.

Cresti, E. 2000. Per una nuova classificazione dell'illocuzione a partire da un corpus di parlato (LABLITA). In E. Burr (ed.), *Atti del VI Convegno internazionale SILFI*. Duisburg.

Firenzuoli, V. 2003. Le forme intonative di valore illocutivo dell'italiano parlato. Analisi sperimentale di um corpus di parlato spontaneo (LABLITA). PhD diss., Università degli Studi di Firenze.

Fónagy, I. 1993. As funções modais da entoação. *Cadernos de Estudos Lingüísticos* 25: 25-65.

Fónagy, I. 2000. *Languages within Language*. Amsterdam: John Benjamins.

[8] We have distinguished here, within the vocal channel, a supra-glottal (voice quality) and a glottal component (prosody).

Fónagy, I. & Bérard, E. 2006. Functions of intonation. In Y. Kawagushi, I. Fónagy & T. Morigushi (eds), *Prosody and Syntax*. Amsterdam: John Benjamins, 19-46.

Gussenhonven, C. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Ladd, D.R. 1978. *The Structure of Intonational Meaning*. Bloomington: Indiana University Press.

Ladd, D. R. 1983. Phonological features of intonational peaks. *Language* 59: 721-759.

Ladd, D.R 2008. *Intonational Phonology*. Cambridge: Cambridge University Press, 2nd. Edition.

Léon, P. 1993. *Précis de Phonostylstique*. Paris: Nathan.

Levitt, E. 1964. The relationship between vocal and facial communication abilities. In J. Davitz (ed.), *The Communication of Emotional Meaning.* 87-100.

McKay, T. & Michael, N. 2010. Propositional Attitude Reports. In N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2010 Edition). http://plato.stanford.edu/archives/ win2010/entries/prop-attitude-reports

Moraes, J. 2008a. The pitch accents in Brazilian Portuguese. Emotion expression in speech acts in Braziliam Portuguese: production and perception. In *Proceedings of Speech Prosody 2008*. Campinas, Brazil.

Moraes, J. 2008b. Atitudes, atos de fala e entoação. Paper presented at the *I Seminário de Pragmática e Prosódia*. Faculdade de Letras, UFRJ.

Moraes, J., Rilliard, A., Mota, B. & Shochi, T. 2010. Multimodal perception and production and of attitudinal meaning in Brazilian Portuguese. In *Proceedings of Speech Prosody 2010*. Chicago.

Moraes, J., Rilliard, A., Erikson D. & Shochi, T. (submitted). Perception of attitudinal meaning in interrogative sentences of Brazilian Portuguese. Submitted to the *XVII International Congress of Phonetic Sciences.* Hong Kong, 2011.

Moraes, J. & Rilliard, A. (forthcoming). The intonation of attitudinal meaning in assertive and interrogative sentences of Brazilian Portuguese.

Osgood, C. & Tzeng, O. 1990. *Language, Meaning, and Culture: The Selected Papers of C. E. Osgood*. New York: Praeger.

Pakosz, M. 1983. Attitudinal judgments in intonation: some evidence for a theory. *Journal of Psycholinguistic Research* 12: 311-326.

Prieto, P. 2003. Teorías Lingüísticas de la Entonación. In P. Prieto (ed.), *Teorías de la Entonación*. Barcelona: Ariel.

Quine, W. 1956. Quantifiers and Propositional Attitudes. *Journal of Philosophy* 53: 177-187.

Reis, C. 2007. A prosódia na expressão de atitudes do locutor no ato de fala. Paper presented at the *I Colóquio Brasileiro de Prosódia da Fala.* UFRJ.

Reis, C. 2010. A expressão da certeza, da incerteza e da dúvida em português brasileiro através de meios prosódicos. Paper presented at the *5th LABLITA International Workshop in Corpus Linguistics and 2nd Brazilian Seminar on Pragmatics and Prosody: Illocution, modality, attitude, information patterning and speech annotation*. UFMG, 2010.

Rilliard, A. 2010. Projet PADE (Prosodie: Accents, Dialectes, Expressivité), Programme Jeunes Chercheuses et Jeunes Chercheurs, Agence Nacional de la Recherche.

Rilliard, A., Shochi, T., Martin, J-C., Erikson, D. & Abergé, V. 2009. Multimodal indices to Japanese and French Prosodically expressed social affects. *Language and Speech* 52: 223-243.

Russel, B. [1918] 1985. *The Philosophy of Logical Atomism*. Peru (Illinois): Open Court.

Scherer, K. 2000. Psychological models of emotion. In J. Borod (ed.), *The Neuropsychology of Emotion*. Oxford: Oxford University Press, 137-162.

Scherer, K.R. & Bänziger, T. 2004. Emotional expression in prosody: a review and an agenda for future research. In *Proceedings of Speech Prosody 2004*. Nara, Japan.

Searle, J. 1969. *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Searle, J. & Vanderveken, D. 1985. *Foundations of Illocutionary Logic*. Cambridge: Cambridge University Press.

Shochi, T., Aubergé, V. & Rilliard, A. 2007. Cross-listening of Japanese, English and French social affect: about universals, false friends and unknown attitudes. In *Proceedings of the XVIth International Congress of Phonetic Sciences*. Saarbrücken.

Stibbard, Richard .2000. Automated extraction of ToBI annotation data from the Reading/Leeds emotional speech corpus. In R. Cowie, E. Douglas-Cowie & M. Schröder (eds), *Proceedings of the ISCA workshop on Speech and Emotion*. 143-148.

't Hart, J., Collier, R. & Cohen, A. 1990. *A Perceptual Study of Intonation: an experimental-phonetic approach*. Cambridge: Cambridge University Press.

Vanderveken, D. 1990. *Meaning and Speech Acts. vol. 1 Principles of Language Use*. Cambridge: Cambridge University Press.

Ward, G. & Hirschberg J. 1988. Intonation and propositional attitude: the pragmatics of L* + H L H%. In *Proceedings of the Fifth Eastern States Conference on Linguistics*. Columbus: Ohio State University Press, 512-522.

Wichmann, A. 2000. The attitudinal effects of prosody, and how they relate to emotion. In R. Cowie, E. Douglas-Cowie & M. Schröder (eds), *Proceedings of the ISCA workshop on Speech and Emotion*. 143-148.

THE DEFINITION OF FOCUS IN LANGUAGE INTO ACT THEORY (LACT)

Emanuela Cresti
LABLITA, University of Florence

1. Premises on Language into Act Theory (LAcT)

1.1 The pragmatic nature of Comment

1.1.1. In LAcT[1] the information structure of the utterance is pragmatically based and it is treated according to the Information Patterning Hypothesis. The starting point of the information patterning (IP)[2] is the accomplishment of the illocutionary force by the specific information unit (IU) named Comment. An utterance can be simple, i.e. compounded of only one Comment IU, and according to the data of C-ORAL-ROM (Cresti & Moneglia 2005; C-ORAL-ROM) nearly 43% of the utterances are simple in romance spoken languages.

Below is an excerpt from a familial conversation in the Italian section: it was taken during a show of old pictures and it is compounded exclusively of simple utterances[3].

---

[1] See Cresti (1987; 2000; 2006), Cresti et al. (2011).

[2] In the literature the terminology regarding the organization of the utterance information goes from *information packaging* (Chafe 1970), to *information structure* (Krifka 2006), to *phrasing* (Gabriel & Lleò 2011), and in our terms is *information patterning* (Cresti 1994).

[3] Our transcription is a version of the CHAT format (McWhinney 1994) integrated with the tagging of terminal and non-terminal prosodic breaks (Moneglia & Cresti 1997); the transcription is orthographic and capital letters are employed only for proper names. Speakers are identified through one asterisk and three capital letters, followed by a colon and one space; each dialogic tour is introduced by the acronym of the speaker and goes on until his silence. Each utterance and each information unit are followed after a space respectively by a double or single slash (//, /). A slash is marked by its informational tag with three capital letters (COM, TOP, ALL, etc..). In a dependent layer, preceded by the percentage symbol, there can be different kinds of information, specifically of that concerning the type of illocution, the information patterning, the situation, the lexicon (%ill, %inf, %sit, %lex). Other transcription conventions regard the diacritique '*&*' as symbolizing a word fragment and '*+*' as symbolizing an interrupted utterance. The diacritiques [/] and [//] represent the phenomena

Mello H., Panunzi A., Raso T. (eds), *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation*

(1) *ELA: o chi l' è questa ?COM 'who is that one?'
%ill: partial question
*LIA: 'un c' indovini //COM '(you) cannot guess'
%ill: invite (to guess)
*MAX: no /COM 'un ci credo /COM no no //PHA ma tu se' te ?COM 'no, (I) can't believe it, no no. But it is really you?'
%ill: [1]expression of disappointment; [2] request of confirmation
*LIA: <no> //COM 'no'
%ill: disconfirmation
*ELA: <no> //COM 'no'
accomp %ill: agreement
*MAX: chi è / Sonia ?COM 'Isn't it Sonia?'
%ill: request of confirmation
*LIA: è la Malvina //COM '(Here) is Malvina'
%ill: presentation
*MAX: mamma <mia> //COM 'mother!'
%ill: expression of disappointment
*LIA: la genovese //COM 'the genovese'
%ill: expression of disdain
*ELA: <ah !>COM
%ill: understanding [ifamcv01]

The simple informational organization of (1) can be easily appreciated, but piece by piece a great variation of communicative actions emerges in so brief an excerpt of conversation. Moreover, the typology of the illocutions (partial question, agreement, presentation, expression of disdain, invite, request of confirmation) appears quite different from types reported in traditional taxonomies, such as for instance the Searlian one (Searle 1969). Actually, during the last decade the analysis carried out by LABLITA has led to the identification of a larger set of about 90 speech act types (Cresti & Firenzuoli 1999; Firenzuoli 2003; Cresti 2006; Moneglia 2011), found empirically. The criteria of their classification imply pragmatic identification and definition, lexical and prosodic correlations, and frequency data.

1.1.2. The pragmatic identification of an illocutionary type is developed during the observation of the corpus, which is analyzed and annotated with respect to information patterning, prosody, lexicon, and syntax. This work allows the recognition and description of similar acts, which can be assimilated on the basis of specific features, despite their idiosyncratic linguistic content. The crucial features

---

of retraction, such as repetitions or reformulations. Overlaps are marked with angular parentheses (<word>).

identifying the nature of a speech act are: the communication channel, the attention, the proxemics between the speakers, intentional features of the process, effects, modifications on the partner, perceptual characters of the referred ontological entity in the pragmatic/cognitive context, the preparatory condition in the speaker, the preparatory condition in the hearer. Moreover it must be underlined that a lot of illocutionary types are performed through prosodic units (PU) of the *root* type ('t Hart et al. 1990); in LABLITA experimental research has identified at least 30 *root* types with idiosyncratic shapes dedicated to the expression of specific illocutions (Firenzuoli 2003). They often constitute the decisive mark for the attribution of an illocutionary type[4].

The schemas in Table 1, 2, 3, and 4 are instances of the pragmatic description of some common illocutionary types.

Table 1.

	Answer	**Order**
Communication channel	open	open
Attention	shared	shared
Proxemic between the speakers	direct interaction	direct interaction
Intentional features of the process	cognitive	operative
Effects	shared information focus	modification of the world
Modifications in the partner	cognitive	operative
Perceptual characters of the referred objects in the pragmatic/cognitive context	no restriction	presence of the referred ontological entity in the context
Preparatory condition in the speaker	question by the hearer	social role and/or pragmatic skill
Preparatory condition in the hearer	expectation	possibility of intervention in the pragmatic situation

[4] We can only mention that in the case of indirect speech acts, according to Grice's terminology (Grice 1975; Searle 1975) their real production often corresponds to the proper prosody of the direct act accomplished. For instance 'can you pass me the salt' is not usually performed with a questioning prosody but with that of a request, sometimes even unkindly.

Table 2.

	Answer	**Conclusion**
Communication channel	open	open
Attention	shared	shared
Proxemic between the speakers	direct interaction	no interaction
Intentional features of the process	cognitive	cognitive
Effects	shared information focus	not shared focus
Modifications in the partner	cognitive	not implied
Perceptual characters of the referred objects in the pragmatic/cognitive context	no restriction	proximal
Preparatory condition in the speaker	question by the hearer	problem in the context
Preparatory condition in the hearer	expectation	no restriction

Table 3.

	Order	**Instruction**
Communication channel	open	open
Attention	shared	shared
Proxemic between the speakers	direct interaction	direct interaction
Intentional features of the process	operative	cognitive
Effects	modification of the world	modification of knowledge and abilities
Modifications in the partner	operative	cognitive
Perceptual characters of the referred objects in the pragmatic/cognitive context	presence of the referred ontological entity in the context	possibility to explore the content
Preparatory condition in the speaker	social role and/or pragmatic skill	knowledge
Preparatory condition in the hearer	possibility of intervention in the pragmatic situation	need of know-how

Table 4.

	Expression of obviousness	**Softening**
Communication channel	open	open
Attention	shared	shared
Proxemic between the speakers	direct interaction	direct interaction
Intentional features of the process	evalutative, expression of a conformity belief	evalutative, expression of caution
Effects	reinforcement of the social link	avoiding clash of opinion
Modifications in the partner	attitudinal (empathy)	attitudinal (agreement)
Perceptual characters of the referred objects in the pragmatic/cognitive context	presence of a shared common ground of values	repetition of the addressee judgment
Preparatory condition in the speaker	presumed same social role	pursuit of agreement
Preparatory condition in the hearer	disposal to the acceptance	opening to negotiation

1.1.3. According to our research it has been possible to identify a provisory and open repertory of speech act types 5, as below in the schema in Table 5.

In conclusion, given that the 43% of the utterances are simple i.e. they are compounded of a unique Comment IU, the structure of their IP is reduced to the expression of one illocutionary force, which is performed through a specific PU root.

[5] The schema is that conceived by Moneglia (2011). A comparison with the repertory proposed by UBLI has been proposed by Cresti (2006).

Table 5.

Representa-tives	**Directives**	**Expressives**	**Rites**	**Refusals**
Concluding	Distal recall – not visible object	Exclamation	Thanks	
Make assertion	Distal recall – visible object	Expression of contrast	Greetings	
Answering	Proximal recall	Expression of obviousness	Apologies	
Commentary	Distal deixis	Softening	Welcome	
Strong assertion	Proximal deixis	Expression of surprise	Congratula-tion	
Identification	Presenting (object / event)	Expression of fear	Wishes	
Verification	Introducing (person)	Expression of relief	Compliments	
Claim	Request information	Expression of uncertainly	Declaration of legal value	
Hypothesis / Supposition	Request of action	Expression of doubt	Condemnation	
Explanation	Order	Expression of certainty	Condolences	
Inference	Total question	Expression of wish	Baptism	
Definition	Partial question	Expression of disbelief	Promise	
Narration	Alternative question	Expression of pitty	Bet	
Describing	Request of confirmation	Irony		
Quotation	Reported speech	Regret		
Objection	Announcing	Complaint		
Confirmation	Advising	Imprecation		
Approval	Warning	Insinuation		
Disapproval	Suggestion	Derision		
Agreement	Proposal	Provocation		
Disagreement	Recommend	Reproaching		
	Invite	Hint		
	Prompt	Encouragement		
	Permit	Assuring		
	Authorize	Threatening		
	Prohibition	Giving up		
	Instruction			

1.2 Comment is necessarily new

1.2.1. LAcT is an extension of the Speech act theory by Austin (1962), which has developed over the years depending on the systematic study of large spoken Italian and romance corpora, and specifically on the observations deriving from the study of the alignment of transcript texts/sound. Evidence emerging from this kind of data led us to propose a pragmatic basis for the information structure of the primary reference entity of speech i.e. utterance, and to recognize prosody as the mandatory mark of its IP.

Following Austin, in LAcT the illocutionary act is conventionally defined and its typology too is conventionally founded, even if it is empirically recognized and characterized such as we have shown in 1.2.. However, a novelty of LAcT is the conception of the perlocutionary act which in spite of being defined as a non conventional intention/effect is defined such as the affective base which is at the origin of the entire speech act (Cresti & Firenzuoli 1999; Cresti 2000). Specifically, just the type of illocution depends on the affective disposition of the speaker toward the addressee; for instance, independently of what should be the content of an utterance, the same mental representation can be turned to the addressee as an order, a polite request, an instruction, a question, an invite, a suggestion, etc., following the kind of relationship occurring between the speakers. The type of the speaker's behavior depends directly on the affect motivating him. The psychic dynamics between speakers is the driving force of speech and it is continuously changing and becoming unpredictable.

This characteristic has a direct consequence on the evaluation of the IP of the utterance, because if the speaker affects govern the speech, which is continuously changing, then the IP results are also unpredictable. Thus, the addressee is always unable to foresee what the speaker's next illocutionary act will be, and as a consequence uncertain about what his own affective and pragmatic reaction should be, too.

What results is always unknown and new in a dialogue and is the accomplishment of the next illocutionary act by the speaker, thus the information about the value and type of the speaker illocution becomes central. Due to the affective origin of the illocution, the Comment IU represents the necessary and most informative part of the utterance[6].

1.2.2. The corpus observation shows that even when the Comment's linguistic content is already present in the dialogue, or it has even been literally said, and in

[6] It must be remembered that Comment is necessarily signaled by prosody, as with every IU, but moreover prosody also specifies its illocutionary type.

conclusion it had to be considered given in some respect, it becomes new if the "old" expression is the Comment of the utterance accomplishing its illocution. This appears with clear evidence from data of spontaneous speech corpora, within which the *third turn principle* occurs with high frequency. The third turn principle is the name for the repetition by the same, and/or by different speakers, of the same, or very similar, linguistic content, in the course of a negotiation. The same expression is performed many times such as Comment IUs with the change of illocutionary values. Actually, even if the linguistic content had to be considered given from a semantic point of view, it becomes new just for the change of its illocutionary performance which is the necessary information expected by the participants of the negotiation.

Below are some Comments with "old" linguistic filling and new illocutionary value.

(2) *LIA: è quella araba //COM '*is that Arabic*'
%ill: assertion
*ELA: araba //COM '*Arabic?*'
%ill: doubt
*ENO: araba /COM sì //PHA '*Arabic, yes*'
%ill: upset confirmation [ifamcv01]

(3) *KAT: o Mingro //COM mi dai una rondella //COM '*Hi Mingro. Can you pass me one rondella?*'
%lex: *rondella* is the name of a round biscuit
%ill: [1] recall; [2] request
*MIC: una rondella ?COM '*one rondella?'*
%ill: surprise question
*KAT: sì //COM '*yes*'
%ill: positive assertion (confirming the previous request)
*MIC: una rondella //COM '*one rondella*'
%ill: answer of confirmation to his previous question [LABcorpus]

(4) *SRE: ma infatti si fa + non per tre //COM '*but indeed it has to be done + not for three (persons)*'
%ill: disconfirmation
*GNA**:** per persona //COM '*for (each) person*'
%ill: hypothesis
*SRE: per persona //COM '*for (each) person*'
%ill: assertion
*GPA: per persona //COM '*for (each) person*'
%ill: confirmation [ifamcv02]

Usually people within the negotiation sequence, before passing to a new subject of conversation, wait for the accomplishment of the third (or last) turn with a final confirmation illocution.

In conclusion, the accomplishment of an illocution performed by an expression, which in this way develops the information function of Comment, makes it the central and new part of the utterance, independent of any feature of its semantic content[7].

1.3 The Topic-Comment relation

1.3.1. If 43% of utterances are simple, the other 57% of utterances correspond to an information pattern (IP) whose origin and center is still represented by the necessary Comment IU, but which can be integrated and supported by other types of IUs.

The schema in the next page summarizes types, definitions and some general features of IUs such as they were identified by experimental work in LABLITA (Figure 1).

1.3.2. LABLITA developed a data base of the IP regarding the informal part of the C-ORAL-ROM Italian session: the IPIC Corpus. IPIC archives the informational annotation of 20835 terminated sequences of informal Italian speech that are characterized with a good or mid acoustic quality. 10,733 of these terminated sequences are compound i.e. the IP of the sequence records at least one IU more than the necessary one of Comment, making up 51% of the total[8]. In accordance with IPIC data, we claim that the primary IP of the utterance is the Topic- Comment pattern, because it represents 23% of compound terminated utterances, in spite of the 8% of the IP Comment-Appendix and 5% of the IP Comment-Parenthesis. These percentages confirm the relevance of the information function of the Topic IU with their quantitative evidence.

Generally speaking, one of the most accepted definitions of the Topic's relation with the Comment (or so-called Focus) has been that it corresponds to a *semantic aboutness* (Chafe 1976). This kind of relation allows that expressions in Topic and those in Comment enter to compound a unique semantic entity, more or less with a propositional size.

[7] The only one condition is that a pure morpheme (article, clitic pronoun, conjunction, preposition) cannot develop an illocutionary function

[8] The general percentage with respect to C-ORAL-ROM data decreases from 57% to 51%, but this can be explained with the difference between the diaphasic composition of the two corpora involved.

Figure 1.

Type of Unit	Name	Tag	Definition	Informational Variation
Textual	Comment	COM	It accomplishes the illocutionary force of the utterance. • It bears a semantic *Focus* (pragmatic salience) • A Comment IU is necessary and sufficient to perform an utterance.	Roughly 90 illocutionary types
	Topic	TOP	It identifies the domain of application for the illocutionary act expressed by the comment, allowing a cognitive reference (relevance) to the speech act. • The Topic let the utterance be displaced from the context (linguistic and extralinguistic). • The Topic necessarily bears a semantic Focus (cognitive prominence)	• Referential objects • Modals • Connectives (according to some practices)
	Appendix of Comment	APC	It integrates the text of the Comment and concludes the utterance, marking an agreement with the addressee, • Appendix never bears a focus	• Repetition • Delayed information • Filler
	Appendix of Topic	APT	It gives a delayed integration of the information given in the Topic adding specification for the addressee.	• Lexical integrations, • Amendments • Repetition of the locution in Topic
	Parenthesis	PAR	it adds to the utterance an information with a meta-linguistic value having *backward* or *forward* scope • Always bears a modal value. • Never bears a focus	• *Modal evaluation* • *Addition of information* • *commentary on the activities by the participants in the communication context* • *Terminological glossa* • *Locutive glossa of reported speech*
	Locutive Introducer	INT	It signals that the subsequent locutive space (simple or patterned) has an unitary *point of view and/or modality*	The locutive space is characterized by spatio-temporal coordinates differing from those of the utterance when it introduces meta-illocutionary comments, such as: • *dicendi* formulas followed by *Reported speech* • *narration* formulas followed by a *Fiction* • *putandi* formulas followed by *Spoken thought* • *presentative* and *deictic* formulas followed

Given the LAcT's pragmatic perspective, the definition of the Topic diverges from this shared assumption, because Topic is defined such as the *application field for the illocutionary force* of the Comment[9]. This notion needs some explications.

The Topic, in order to be an adequate field for the application of the illocutionary force, must play the role of a contextual prominence within the utterance. It must be underlined that the fact that the Topic represents a contextual prominence through linguistic devices, does not mean that Topic is a piece of context (linguistic, semantic, pragmatic) directly subsumed in the utterance.

In the case that the Topic is missing, it is assumed that the Comment has to refer to the context and apply its illocutionary force to it in a so-called deictic way, like the most part of literature claims lacking a pragmatic perspective. Actually we think that the Comment refers to the context not simply in a deictic way, but according to specific aspects among which is its illocutionary type. If by chance the force of the Comment is an order, it pursues a certain intervention in the world by the addressee.

(5) shut the door //COM
%ill: order

For what concerns the situation in (5), if the speakers are in a room with a door and a window open, the utterance specifies what operation the addressee is expected to comply with, including the object to be considered as having the right pragmatic prominence involved. The entire Verbal phrase, with its syntactic relation Verb-Object, is the expression bearing the force, so that the type of force performed elicits in some sense also a part of context. Anyway, we disagree on the assumption that this reference is deictic, if the meaning of the term is restricted to a "pointing function", because it is the entire meaning of the phrasal expression that refers to the context and it is not empty like deictical expressions are. The only case in which the reference can be considered deictic occurs in (5a) within the following sequence. Let us compare (5), (5a) and (5b):

(5) shut the door //COM
%ill: order
%sit: the addressee recognizes the order in its whole, including the semantic denotation of its object of intervention

[9] This definition can appear unexpected in the linguistic tradition, because if the definition of Comment IU in terms of illocutionary force is rare and not really exploited in the literature (Jakobs 1984), the consequent definition of Topic in terms of its field of application seems to be a real novelty

(5a) shut it //COM
%ill: order
%sit: the addressee recognizes the order, but he must look for the adequate pragmatic prominence in the context, corresponding to "it", and in this case the order refers deictically to the context.

Thus the fact that the illocutionary force applies to the context does not mean that it refers necessarily to it with a pointing operation, as it happens when a deixis is accomplished finding the content of the act in the context.

Moreover the situation is different in (5b), whose IP is compounded with a Topic:

(5b) the door /TOP shut it //COM
%ill: order
%sit: the addressee recognizes the order, but he has been supplied with the information relevant to the appropriate contextual prominence to take into consideration for his intervention, through the linguistic expression of Topic.

In this regard, in (5b) the 'door', functioning as Topic, is not the semantic nor the syntactic object of the verb 'shut', functioning as Comment as it does in (5), but it is the *linguistic representation of a contextual prominence* to which it is expected the order will attend. We define this specific relation in informational terms like a function of *pragmatic aboutness*, that is developed by the Topic with respect to the Comment.

But the most relevant data deriving from the observation of spoken corpora is the discovery of the variety of illocutionary types and the frequency of some types which are even ignored by literature, so that if by chance the force is an *expression of obviousness*, like in (6), then the model proposed for (5) must also be changed.

(6) (if) you smoke forty cigarettes…COM (you'll probably fall ill)
%ill: expression of obviousness

Actually, given that the goal of this expressive force is the sharing of a common conclusion with the addressee, in this case the conclusion has to be conceived and suggested to the addressee by the same Comment. A force like the expression of obviousness seems to do more than elicit whatever part of the context had better be at the same conception and origin of a conclusion that could in principle be shared. This conclusion was not already available even in the cognitive universe, but it is derived from the expression of obviousness with regard to certain content. So it is impossible to claim that there is a kind of deixis to something already present; in some respects the expressive force creates its same domain of application which is added to the context by the speech act. So given that even the relation of Comment

with the context cannot be defined in principle and cannot always be deictic, the relation of Comment with Topic cannot be defined any more like a deictic operation. Topic, as the linguistic representation of a pragmatic prominence, to which the Comment force must apply, seems to be more an entity "suggested" by the latter.

The way the Comment refers to the Topic varies in accordance with its force and its linguistic content and step by step, the Topic must be able to play the role of a pragmatic prominence, adequate to the type of force accomplished by the Comment, doing so with linguistic devices.

In conclusion, in spite of the traditional definition of the Topic's relation with the Comment in terms of a semantic aboutness, leading to a propositional entity, in the pragmatic perspective of LAcT the relation of Topic with Comment can be synthesized in term of a pragmatic aboutness, whose goal is the representation of a linguistic domain adequate for the application of the illocutionary force.

1.3.3. A consequence of the pragmatic function of Topic, as the field of application of the illocutionary force, is the full satisfaction of the request proposed by Hockett (1958) that a Topic allows the *displacement* of the Comment from the context[10].

It must be observed, that, perhaps, it has not been considered enough how this assumption reverses a general and shared perspective on Topic. Actually if a Topic must displace the Comment from the context, it means that it cannot be a "piece of context" which takes part in the utterance. For instance, a common definition of Topic is founded on its semantic *oldness* or *giveness*, claiming that this semantic feature belongs to an expression if its denotation is already present in the Common Ground (CG)[11], or at least shows a certain degree of presence in it. But if the nature of the Topic were that its denotation had to be available, something already present in the text, or in the discourse universe, or in the encyclopedia, it should mean that Topic is derived and anyway dependent on the context. For this reason Topic did link the Comment and finally all the utterance to the context and not in the contrary allow its displacement from it.

Also the acknowledged definition of Topic given by Reinhart (1982) led to the same conclusion:

> The notion of 'topic' comes with a complementary part called 'comment.' [..] New information is not just added to the Common Ground (CG) content in form of unstructured propositions, but is rather associated with entities, just like information in a file card system is associated with file cards that bear a particular heading.

[10] We note that it was Hockett that introduced in the USA the terminology of Topic-Comment, translating the Praguian couple *theme-rheme* (Sornicola & Svoboda 1989) , but in effect opening the possibility of a new meaning of this terminology.

[11] We will come back to the concept of Common Ground in the third paragraph, for the moment it can be interpreted such as a variant for the concept of context.

Even if Reinhardt said that a Topic cannot be existent without a Comment, and in some sense let understand a dependency of Topic on Comment, she assumed that the new information carried by the Comment is inserted in the CG not in a free and uncontrolled way but through the index of the Topic, such as it were a cue already being part of the CG. In Reinhardt's view Topic gives the right entrance in the CG to the Comment, in that way linking properly the Comment to the CG. So also in this perspective no displacement function could be possible.

Even if we would ignore research arguing in favor of Topic with new informative content (Berrruto 1985), in all cases we cannot ignore corpus data. They record many occurrences of Topic (roughly 15%), whose denotation is new and appears for the first time in the discourse. Below a piece of conversation is reported:

(7) *NIC: # cosa succede ?COM '*what is going on ?'*
*CEC: eh il colore del palco /SCA é una brutta decisione //COM [#] hhh allora /INP l' hai <trovata> ?COM '*the color of the stage is a hard decision. So, did you find it?*
*NIC: [<] <questa> no /COM aspetta //CNT '*this no, wait.'*
*CEC: **di là /TOP gli acidi /TOP tutto pronto ?COM** '*(for what concerns) in the other room, (for what concerns) the acids, everything ready?*'
*NIC: questo dovrebbe essere //COM # cos' è ?COM un xxx <quante [/1] quante sono> + '*this should be right. What is it ? How many ? How many are ?'* [ifamdl17]

In the fourth turn of (7), all of sudden the speaker CEC creates a space of application 'di là' (*in the other room*) and an object of application 'gli acidi' *(the acids*) which do function such as two adequate fields for the question force (*everything ready?*). But both are new and are unexpected in that excerpt of dialogue, actually they are motivated only by a CEC's snap anxiety. She is thinking and speaking about a matter which is suddenly emerging for herself, but which is totally absent in the shared situation , so that the other speaker NIC does not even answer to the question which he probably does not understand. But it does not mean that the utterance with two new Topics has not been performed, and it is a less exceptional instance than could be imagined.

Data records a significant percentage of Topics that cannot be defined according to a general feature of giveness, and this is a sufficient proof that giveness cannot be the crucial semantic feature identifying Topic. The fact that often the semantic content of Topic corresponds to persons, objects, events, arguments, judgments, times, already present in the situation, is not the crucial aspect.

Anyway the fact that the majority of Topic semantics can be characterized by a feature of giveness asks for an explication. It can depend on many reasons: in the case of a more discursive text, with descriptions and argumentation, the given content of Topic is functional for the textual architecture of the whole. There are rhetoric reasons asking for reprises and kinds of anadiplosis, but these regard a different level of organization from the basic one of the utterance. General

characters of argumentation cannot be confused with the information structure of utterances, which exists also in unprepared texts.

Moreover the diamesic nature of speech, which must develop itself within a shared situation by the speakers, pushes toward a choice of prominences which can be easily recovered by the hearer. But this has not to create any misunderstanding, because it cannot be forgiven that the crucial character of language is to be about the reality, but at the same time to be independent from it. Context, in a large conception including the human dynamics, constitutes always the stimulus and the input for the speech act but context does not determine language which is a human creation depending on thought and affected by each speaker. There is a solution, a jump, between the context stimulus and the internal, mental and affective reaction. Language is not determined by the context, even if it is always about the context, language is a human creation.

The assumption that a Topic is a linguistic representation of a possible prominence, adequate to the force of the Comment, means that it is determined by the argument of the Comment, by its specific illocution and semantic content, in spite of any character of giveness or novelty. A Topic-Comment IP is self standing, because Topic displaces Comment from the context, and so doing allows reaching one fundamental task of the language: to be free from the world. Below are some examples where the Topic introduces a new argument in the chat, being the field for a total question as in (8), makes a deictical reference, being the field of an expression of intention in (9), and starts the presentation of a work procedure with a generic description in (10).

(8) *EST: riparlando della Pina /TOP hai visto come si veste ?COM '*talking again about Pina, have you seen how she is dressed?'*
%ill: total question [ifamdl15]

(9) *PRO: questo /TOP lo tolgo da qua / COM che non è il posto //APC *'this, (I) take it from here, that it is not its place*'
%ill: expression of intention [ipubdl04]

(10) *ART: forme di borse /TOP essenzialmente /TOP sono due //COM '*the bag forms, actually, here are two (types)*'
%ill: presenting [ifamdl04]

In the previous examples three different semantic types of Topic are shown, but their result is the same with the independence of the utterance from the context.

In Conclusion in LAcT the IP is pragmatically based and its primary pattern corresponds to the accomplishment of an illocutionary force by the Comment, which is therefore new, and to its field of application by a Topic, playing the role of an

adequate prominence, according to a pragmatic aboutness relation. Doing so, Topic allows the displacement of the Comment from the context.

2. Semantic Consequences of the Topic-Comment IP

The conception of Topic-Comment information relation as a pragmatic aboutness has some semantic implications[12].

2.1 The Topic-Comment IP is not a Predication

2.1.1. The Topic- Comment IP does not correspond to a semantic relation of Predication, that is, an utterance compounded of a Topic- Comment is not a Proposition: the Topic is not the Subject of the Proposition and the Comment is not its Predicate[13].

Considering the following examples it is possible to verify our assumption:

(11) *VER: le **mele** /TOP fatte a **cigno**[14] //COM '*(for what regards) the apples, (the right shape should be) like a swan*'
%ill: expression of obviousness [ifamdl14]

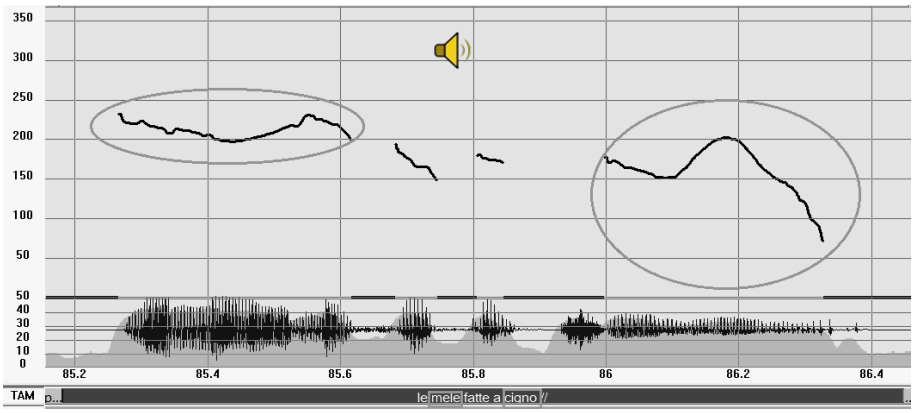


Figure 2.

[12] The argument has been developed in Cresti & Moneglia 2010.
[13] See Li (1976).
[14] The bold character marks the Focus.

(12) *GAB: poi il **barocco** /TOP può non **piacere** //COM '*then (for what concerns) the baroque style, somebody can not enjoy it*'
%ill: observation [ifamcv17]

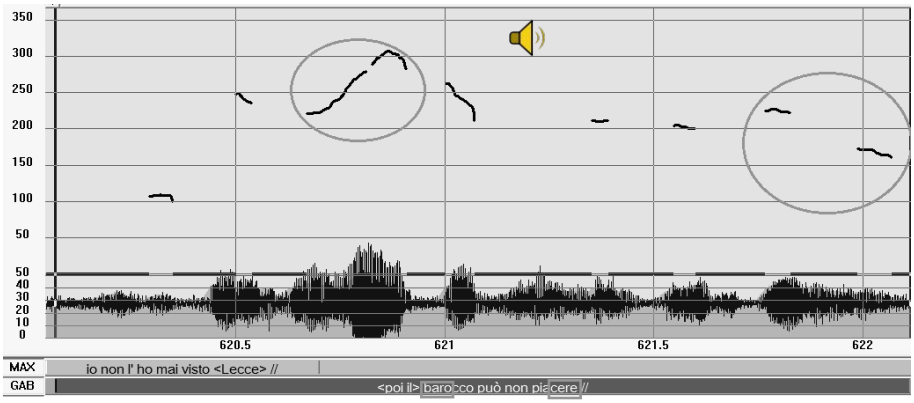


Figure 3.

A bare orthographic transcription of the sound of (11) and (12) will suggest a NP interpretation of (11), *the apples swan like*, or a Sentence interpretation of (12), *the baroque (style) cannot enjoy (somebody)*. But if their prosodic performance is considered, it marks in a necessary way the information role of the linguistic material of each IU, indeed the Topic is performed through a *prefix* PU and the Comment through a *root* PU. Only if the prosodic patterning of the utterance is ignored, the interpretation leads to the previous syntactic conclusions, which do not correspond in any way to the utterance performed.

The development of one information function bounds all the expressions cooperating toward the task in the same semantic domain, but at the same time the function isolates them from expressions concurring to perform a different function. When the speaker puts in action some linguistic material with a certain information function, he behaves in a way pragmatically motivated and his fundamental input is an affect toward the addressee; this activity belongs to the illocutionary act. When the speaker performs a syntactic configuration and a semantic composition, he develops a cognitive and computational activity which belongs to the locutionary act. Even if the illocution and the locution are simultaneous in the performance of the same speech act, they concern different faculties. Moreover without an unconscious motion (perlocution) it is impossible to speak and the informational pragmatic program, affectively directed, dominates the locutionary one.

It is true that what is performed with a different information function can be interpreted *a posteriori* by the addressee such as with a syntactic configuration or a semantic proposition, but it is not what the speaker has put in action: it does not

correspond to his behavior. LAcT distinguishes the two perspectives: what is performed by the speaker and put in action on the basis of his affect and what the hearer's reconstruction and interpretation can be. Considering them together is a theoretical mass besides the fact that it does not correspond to the reality.

The linguistic material of Topic and Comment is not bound either by syntactic and semantic relations across the IU boundaries, signalled by prosody, because each chunk is devoted to the accomplishment of one specific information function. The information patterning is ruled within the illocutionary act and dominates the locutionary structure: the information units conceived for the accomplishment of a certain information function identifies the linguistic unit like a local syntactic configuration and a semantic island.

2.1.2. In (11) the illocutionary force of obviousness accomplished by the Comment (swan like) is applied to the prominence (the apple), proposed like the common decoration of the table for a party, and represented by the Topic through linguistic devices.

(11) *VER: le **mele** /TOP fatte a **cigno** //COM '*(for what regards) the apples, (the right shape should be) like a swan*'
%ill: expression of obviousness' [ifamdl14]

In (12) the illocutionary force of observation accomplished by the Comment (*somebody) can not enjoy it*' is applied to the prominence, (*the baroque style)*, such as a figurative style chosen by the speaker like an example, represented by the Topic through linguistic devices.

(12) *GAB: poi il **barocco** /TOP può non **piacere** //COM '*then (for what concerns) the baroque style, (somebody) can not enjoy it*'
%ill: observation [ifamcv17]

In (11) the relation between the NP in Topic and the AdjP in Comment is an information relation of pragmatic aboutness and does not correspond to a syntactic structure of NP:

the Adj does not modify the N in Topic,
the N is not the head of the whole NP

In (12) the relation between the NP in Topic and the VP in Comment, even if in Italian the Noun records a morphologic concordance with the Verb in Comment, is an information relation of pragmatic aboutness and does not correspond to a syntactic structure of sentence:

the VP does not develop a function of predication with the NP in Topic
the NP is not its Subject

No syntactic structure of attribution or predication is in action between constituents behaving as Topic and Comment. From a syntactic point of view they are always *anacholuta*, and behave such as semantic islands.

The traditional term used by rhetoric for this kind of relation is anacoluthon, and it denotes expressions clearly bound within a same broader semantic entity, but lacking any syntactic link[15]. The following examples can be overtly considered instances of anacoluthon because of their syntactic composition: in (13) there is a prepositional phrase with temporal meaning and an adverbial negative phrase, in (14) an adjective phrase and a verbal phrase.

(13) *SAB: per ora /TOP no //COM ‘*until now, no (it does not)*’
%ill: constatation [ipubdl03]

(14) *APR: mensile /TOP costa un po’ di più //COM ‘*monthly, it costs a little more’*
%ill: explication [LABcorpus]

It can be noticed that in order to receive a correct interpretation in their oral version, anacolutha strictly require a prefix-root prosodic pattern, otherwise they will be meaningless as constituents prosodically integrated, i.e. linearized, within the same PU. They cannot be directly mapped onto a well-formed compositional structure, since they are anacoluthon both from a syntactic and semantic point of view.

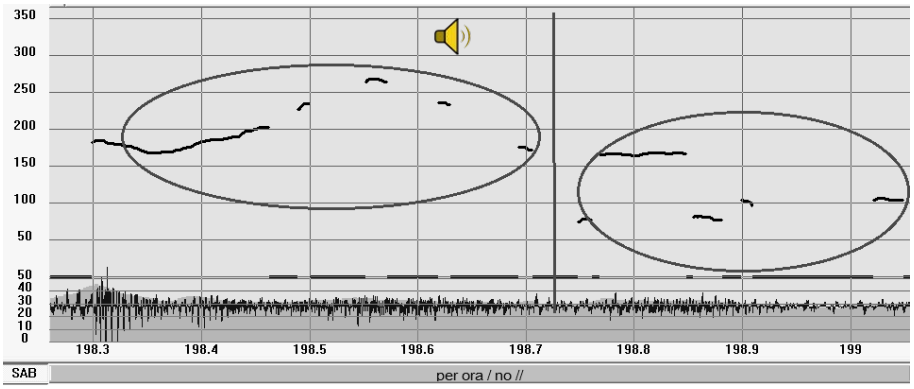


Figure 4.

[15] The term has passed from rhetoric to syntax, where it is currently applied.

Actually, when the speaker does perform full syntactic phrases and sentences, he behaves within the locutionary act with regard to constituents identified for a certain information function, and he marks the syntactic link through a phonetic integration that is called *linearization*. A linearized constituent is performed through only one prosodic unit (PU) recording one major perceptual prominence. In this case it corresponds in its whole to: a local syntactic configuration and a semantic compositional entity. In speech, linguistic expressions participate in the same syntactic configuration and compound the same semantic domain only if they are linearized from a phonetic and prosodic point of view . The entire syntactic and semantic configuration will be devoted to the explication of whatever information function with regard to the illocutionary act. It means that they are united by the development of the same information role and constitute one IU[16].

In conclusion : the Topic is not the Subject of a Proposition and the Comment is not its Predicate. In order that a Subject and a Predicate may develop a semantic relation of predication composing a Proposition, they must be linearized in speech performance within an unique PU. A Topic-Comment pattern is based on a pragmatic aboutness relation within an utterance and their respective IUs must be performed through one *prefix* PU and one *root* PU, with a specific illocutionary value.

2.2 The semantic domain of Topic is idiosyncratic

2.2.1. The second Semantic consequence of the Topic-Comment relation defined as a pragmatic aboutness is that there are restrictions on the semantics of Topic (Cresti & Moneglia 2010). These semantic conditions are not easy to discover and they can be better identified in a contrastive perspective with those of Subject. This data can be obtained only on the basis of corpus based research. The semantic and morpho-syntactic features of expressions occurring in Topic IU and those developing a Subject function actually diverge and record only a limited intersection (Signorini 2005).

Clauses, VPs, Quality Adjectives, Adverbs, appropriately performed through a *prefix* PU, have been found developing a Topic information function, while they cannot be evidently the Subject of a Sentence. Below are some examples

(10) *ART: forme di borse /TOP essenzialmente /TOP sono due //COM ‘*the bag forms, actually, here are two (types)*’
%ill: presentation [ifamdl04]

[16] The comparison between same lexical expressions performed as a Topic-Comment pattern and as a linearized constituent is presented in Cresti & Moneglia 2010, on the basis of experimental works.

(15) *NIC: se ce n' è ancora /TOP uno sì //COM '*if there are some more, one yes'*
%ill: acceptance [ifamdl17]

(16) *LIA: che si doveva fa' perdonare /TOP non l' ho mai voluto sapere //COM '*that she had to be forgiven, (I) did not want to know it*'
%ill: disagreement [ifamcv01]

(8) *EST: riparlando della Pina /TOP hai visto come si veste ?COM '*talking about Pina, have you seen how she is dressed?'*
%ill: total question [ifamdl15]

(14) *APR: mensile /TOP costa un po' di più //COM '*monthly, it costs a little more'*
%ill: information [LABcorpus]

Generally speaking, the semantic domain of the Topic, that must represent a prominence adequate to the illocutionary force of the Comment, seems larger than that of the Subject, because it may include events and properties which on the contrary cannot be employed for the syntactic role of Subject.

2.2.2. The corpus analysis allows us also to verify that *Anaphoric personal Pronouns, Indefinite pronouns* (at least some), *Negative NP* and *existential non generic NP* never occur with a Topic function, nor can they be performed through a *prefix* UP, while these kinds of expressions are possible Subjects and can occur linearized with a Predicate.

Below are some laboratory examples, where the same lexical sequences have been performed such as either sentences in (17), (18), (19) or Topic-Comment patterns (17a), (18a), (19a)[17]. It must be noted that the attempts to make speakers perform the Topic-Comment examples, with their corresponding prosodic contours, were difficult to utter and led to odd results, so that their evaluations by independent native speakers were deemed unnatural.

(17) esso viene risolto //COM '*it is solved'*

(17a) *esso /TOP viene risolto//COM *'*it, is solved'*
%ill: lesson

(18) nessuno si muova //COM '*nobody move*'

(18a) *nessuno /TOP si muova //COM '*nobody, move*'
%ill: enjoinment

[17] It must be remembered that Italian is a PRO-drop language and sentences without Subject are acceptable.

(19) una signora si è risentita //COM ‘*a lady took offence’*

(19a) *una signora /TOP si è risentita //COM *‘*a lady, took offence’*
%ill: narration

Then the semantic domain of Topic seems not only larger than that of Subject , as with our generic intuition, but also narrower, and in conclusion they must be distinguished from one another.

The semantic restriction on Topic can be explained considering that pure anaphora, full negative pronouns, and undetermined existential individual entities do not allow one to substitute a contextual prominence, because they cannot by themselves represent an adequate reference in a proper sense. It must be remembered that Topic ensures the displacement of the Comment from the context, and on the contrary a pure anaphoric expression needs a semantic antecedent to be interpreted, in that way it cannot properly develop the function. Moreover, a total negative expression cannot represent any pragmatic or cognitive domain by itself for evident semantic reasons, and the semantics of a person or an object provided by a kind of denotation, supposed existent but not being identifiable, too cannot constitute the representation of an adequate reference. We claim that none offer a *representative* image by themselves[18].

Then, if the fact that the semantic conditions necessary for representing a pragmatic domain are missing, leading to the impossibility of developing the function of Topic, it means that this condition constitutes the crucial aspect. And it is not a question of novelty or oldness, but a question of *representativeness*.

Moreover, it can be noticed that the fact that expressions like these, lacking representativeness, impede the development of the Topic function, is also a proof that Topic is a semantic island. If they could develop syntactic and semantic relations with the linguistic expressions in Comment no restriction will emerge, as it happens with Subjects. This is shown by simply employing the same expressions with the Subject function. From one side representativeness is not a semantic condition for a Subject, which participates in the propositional composition, and on the other side it is the semantic condition for a Topic, which is a semantic island.

In conclusion, the semantic domain of Topic is idiosyncratic, in particular it is different from that of Subject: larger for the occurrence of events and properties (adjectives and adverbs), and narrower for the absence of negative and indefinite

[18] The employment of this adjective is not due to its first and more common meaning “a person or thing enough like the others in its class or kind to serve as an example or type”, but specifically to “acting or speaking in the place or on behalf of another or others”, which is maintained also in one sense of the derived noun “a person duly authorized to act or speak for another or others”. The Topic function matches with this last meaning because of its “behaving in the place of context”.

individual entities. The semantic condition allowing the accomplishment of its function is *representativeness* and not *giveness*, as traditional frameworks assume.

3. The definition of Focus

3.1 Premises

3.1.1. The first attempts to study and explain the structure of spoken language and its information organization date back from the middle of 1800. It can be cited, as maybe the first research, with concepts such as *point de départ* and *but du discours* by Weill (1844), and *psychological subject* and *psychological predicate* by Gabelenz (1891). Jumping to the second half of 1900, we can remember the most relevant frameworks with the translation of the Praguian concepts of *theme* and *rheme*, imported into the USA with terms like *topic* and *comment* by Hockett (1958; Chafe 1970; Chafe 1976; Gundel 1977) and transformed into *topic* and *focus* (Chomsky 1971; Jackendoff 1972), as well as other approaches proposing *given* and *new* (Halliday 1976a; Halliday 1976b), and *frame* and *center* (Lambrecht 1994).

The structure of information developed by LAcT departs from the track of traditional assumptionsin relation to one particular feature: they do not consider the pragmatic origin of information, ignoring the illocutionary definition of Comment. Moreover they also share two other aspects diverging from LAcT: the semantic nature of Focus, which is substantially identified on the basis of its novelty with respect to context and represents the only key to explaining the information structure, and the fact that Topic derives from the context[19]. In that way the entire information organization of the utterance results are conditioned by the context.

The reason for these differences is that no distinction is foreseen between different activities (illocution and locution) accomplished by the speaker simultaneously, but which diverge in their nature (affective and pragmatic vs cognitive). Given the lack of the illocutionary notion of Comment, in the literature there is no distinction between the semantic concept of Focus and the pragmatic one of Comment. On the contrary, in LAcT the origin of information is the action of the speaker, because of its affective nature which is continuously changing, and which is realized by the Comment. Focus and Comment are concepts belonging to different acts: locutive and illocutive.

[19] If a large meaning of the term *context* is accepted, indeed, Topic can be explained within its extension to cognition (an already present representation in the speaker's mind, or a logic presupposition with regard to the next assertion), or to semantics (a denotation already present in the universe), or to discourse (an element already in the dialogue background).

Traditional semantic definitions foresee that Focus represents the "most important" or "new" information in an utterance. But importance is a vague aspect and can hardly be verified, because, for instance, what is the most important information in an utterance? Can the information of Topic be considered by chance not important? By consequence, is the Topic automatically excluded from the possibility of recording a Focus?

As regards the feature of novelty, on the basis of our corpus data we have already shown that a Comment can record old semantic content from a contextual point of view (becoming new for the illocutionary accomplishment), and that a Topic can record new semantic content (with the only condition being representativeness). If a Topic can be new and a Comment can be old, are importance and novelty opposing values? These questions don't seem to have clear solutions.

Moreover, if spoken language corpus data and prosodic features are considered, it becomes even more difficult to conserve a semantic frame of explication of information structure depending on Context.

3.2 The model of Common Ground

3.2.1. Some acknowledged research on information structure employ the concept of Common Ground (CG) in the place of that of context[20]. The concept has actually been formulated by Stalnaker (1974) and it can be described as «a way to model the information that is mutually known to be shared, which is continuously modified in the course of communication».

CG is a central assumption in the most part of recent theories interested in Focus. It could seem that its conception, such as a kind of context no more still and idle, in some sense matches with what we have been proposing about the information structure of utterance, depending on the continuous change of speaker's affects. But it does not, mostly due to two points: the already cited lack of distinction between Comment and Focus, and the active knowledge of each speaker at the time[21].

The second point derives to some extent from the first one, because if a pragmatic perspective is adopted, any "mutually shared information" cannot exist. All the time that a speech act is performed, it enters the context, changing it, which

[20] By us temporally equalized with that.

[21] The first point leads to the lack of a distinction between the illocutionary pragmatic activity, which is the origin of information and rules its organization, and the semantic level of the locutionary activity, which finds its border/boundaries inside the previous information organization. That is the reason for the always elusive and shifting semantic definitions of information concepts, like Focus/Comment and Topic. Pragmatic definitions of information functions on the contrary are steady and verified on corpora.

becomes the new context. But anyway and always context is "endless" and rich with all kinds of possible inputs available to the hearer. The fact that the context is real, does not mean that it is an independent entity, knowable in its whole as a logic universe. Everybody knows it subjectively, following his mood and giving attention to what is interesting for his own attitude in that moment[22]. There is no mandatory information prominences in the context, but only those inputs which are prominent for the speaker's attention in that moment. Moreover, as we have already said, there is no determination from contextual inputs to the speech act performed, because of the internal affective and mental origin of the latter. The speaker's next speech act is unforeseeable despite of every kind of contextual prominence. Mutually shared information could exist only in a platonic semantic or logic context existing outside of the speakers and in spite of their living actions. CG is a hypothetical semantic structure, existing for itself and in some sense being transcendent, even if it is foreseen that it can change. Only because of its presumed nature is it possible to imagine that it does condition the semantics of Focus.

3.2.2. In some sense a more concrete definition of Focus seems to be given within the framework of Alternative Semantics (Rooth 1992; Krifka 2006). The vague features of importance and novelty are supposedly specified, because

> Focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions. […]This distinction is relevant for information packaging, as the CG changes continuously, and information has to be packaged corresponding to the CG at the point at which it is uttered.

This assumption could seem reasonable and of use, but the claim that information can be packaged «corresponding to the CG at the point at which it is uttered» seems to lead again to a semantic dependence of the information structure on the context. It means that some specific objective features in the context, identifying a point in the CG, condition Focus, given that it still remains a semantic entity. No intuition emerges relative to the accomplishment of an illocution, which is a totally subjective act, and the accomplishment is thus not determined by the context.

Advancing in this line, Krifka (2001) explicates that the prominent use of Focus is the identification of context-questions in answers:

> The idea is that the meaning of a question identifies a set of alternative propositions, the answer picks out one of these, the Focus within the answer signals the alternative propositions inherent in the question.

[22] See example (7).

It is admitted by the same Krifka that the idea is not new, because for instance already neogrammarian scholars (Paul 1880) proponed this perspective, and we add that it has been drawn on also by Bally (1932): Topic/ theme has been explained like a question subsumed inside an utterance, whose rheme is answering to it. In substance, following the Alternative Semantics, the core of an assertion, i.e. the part adding a novelty to the CG, should be the answer chosen by the speaker among the possible ones, given a certain open question in the CG, that may be optionally reported in the theme/Topic. But how to identify the right question in the CG?

Krifka proposes a general distinction about CG content and CG management in parallel with a distinction of a semantic vs. a pragmatic use of Focus, but the last one, which in principle could match with a pragmatic perspective, in our opinion still remains trapped in a platonic semantic universe. Considering acts like question and answer in Krifka's perspective is a purely verbal mention and does not imply any real took in charge of speaker's activity.

> The pragmatic use of Focus is to highlight the part of the answer that corresponds to the wh-part of a constituent question . […] A question changes the CG in such a way as to indicate the communicative goal of the questioner. […] This effect can be modeled by interpreting a question as a set of propositions, each being the denotation of a congruent answer. […] The answer identifies one of these propositions and adds it to the CG content. […] The Focus within the answer signals the alternative propositions inherent in the question.

A relevant extension of the question-answer model is due to theories assuming that a *coherent discourse* is structured by implicit questions (van Kuppervelt 1994, Büring 2003) and by Focus on the answers. The concept of implicit questions foresees that context is characterized by whatever types of features that can constitute or suggest by their selves questions for the addressee.

> Focus allows to accommodate the meaning of questions that are not overtly expressed and generally speaking to accommodate the CG management. [..] All cases of so-called "presentational" or "information" Focus, which is presumed expresses the most important part of utterance, can be subsumed under the use of alternative to indicate covert questions suggested by the context.

In this sense the activity of speech is reduced to answer in a coherent way the questions suggested by the world and the operation could be reduced to a logic schema. The semantic question-answer model transforms the context into an open variable and assumes its satisfaction in the answer, ensuring a result which is characterized by a propositional form. Any pragmatic value of the utterance is not even hypothesized, and this ends the claim of equivalence between utterance and

proposition and to allow the analysis of the former in the semantic terms of the latter.

The previous assumption in spite of its presumed formal solution could still be accused of vagueness, but more exactly it is vacuous. From a theoretical point of view, indeed, it can be always possible to imagine and reconstruct *a posteriori* a feature in the context to justify an answer, i.e. the semantic content of an utterance with a certain Focus. In accordance with LAcT the speech act is not foreseeable, but even starting from a different perspective, given that the aspects of the context are endless, how to identify *a priori* what are the mandatory features in the context at the origin of the possible answer? It could never be proved and verified what is the mandatory aspect of the input and every chosen question in the context can be considered the right one. It means that there is no predictive value in that assumption, whose formal rigor is empty.

But what is most relevant according to our perspective is that corpus data supports the fact that the real spontaneous spoken activity does not occur in this way[23]. The framework of alternative semantics, defining the pragmatic use of Focus as the point marking an alternative in an answer to an overt or covert CG question, does not seem adequate to explain corpus data. Analyzing the stretch of whatever spontaneous dialogue will highlight the impossibility to carry on the discovery of elements to be considered at the origin of covert questions in the CG, so that they are the adequate input for the speech behavior.

3.2.3. Coming back to our example (1), the showing of old pictures seems the perfect situation to find covert questions in the context, allowing the emergence in the dialogue of answers with the proper Foci. But how to justify the first turn reported, which is a partial question?

(1) *ELA: o chi l' è **questa** ?COM '*who is that one?'*
%ill: partial question

How can the unrecognizable image of somebody be considered the covert question giving origin to a partial question? This brings to light a fundamental matter : more than 40% of illocutions accomplished in the spontaneous speech are not assertive. How is it possible to pass from an objective prominence, which could be chosen like the input generating the covert question in the CG, to a speech act which is not assertive but which had to anyway be an answer to that covert question? How is it

[23] Probably research carried out on map-task data, or call center conversations, or other kinds of ruled spoken exchange will allow a different perspective, because within a shared and limited context the task of their participants is exactly that of posing questions and giving appropriate answers. But even in these instances it is easy to find continuous counter-examples.

possible in the specific case to pass from the unrecognizable person in the picture, assumed as the mandatory input, to the action of the partial question, a directive act toward another participant in the conversation? Why has the speaker 'ELA', under this input, chosen to make a partial question and not by chance to claim that she does not recognize the person, or express that she is bored, or to be silent and pass over to another picture?

Then the speaker in the second turn - 'LIA' - had to answer to the overt question of ELA, for this presenting the canonical situation "overt question of the speaker and determined answer of the addressee", but it does not turn out like this. There is a mismatch, because LIA does not answer and poses a riddle:

(20) *LIA: 'un c' **indovini** //COM '*(you) cannot guess*'
%ill: invite (to guess)

What should be the presupposed contextual question input in this case? It could seem to be the will of provocation by the same speaker LIA. How can we consider that this subjective, internal attitude takes part in the context? Personal emotions, attitudes, feelings, can be considered within the Context? How many are there, how do we classify them, and finally, how to prove that one of them is the right one?

Then the first utterance in the third turn is made by a new participant 'MAX', who displays his disappointment, because he thought to have recognized the person in the picture, even if actually he was wrong.

(21) *MAX: **no** /CMM '**un ci credo** /CMM no no //PHA ma tu **se**' **te** ?COM '*no, (I) can't believe it, no no. But it is really you?*' [24]
%ill: [1]expression of disappointment; [2] request of confirmation

Could his false recognition be considered the covert question in the context at the origin of the expression of his disappointment? And given that he keeps on in the second utterance of his turn with a request of confirmation to LIA: could his own disappointment derived from his wrong recognition be considered the covert question in the context at the origin of his request of confirmation?

Even if we go back to example (7), the Comment with a question force does not seem to find any contextual motivation for a snap of anxiety suddenly emerging in the speaker. What could be the proper input for the anxiety which is at the origin of the question?

[24] The tag CMM marks the chain of two COM belonging to the same rhetoric entity. See Cresti et al. (2011) and Cresti (forthcoming).

(7) *CEC: **di là** /TOP **gli acidi** /TOP tutto **pronto** ?COM ‘*(for what concerns)in the other room, (for what concerns) the acids, everything ready?*’ [ifamdl17]

And it must be noticed that no answer is given also to this question by his addressee.

Even reflecting on the sense of third turn principle, how is it possible to reduce the change of illocutions on the same expressions by speakers in the same shared situation because of different covert questions in the context? For instance in (3), the sequence of negotiation, passes by from a request of the first speaker, to a surprise question by the second speaker, to a positive assertion by the first speaker confirming his previous request, to a confirmation by the second speaker to his previous surprise question.

(3) *KAT: o **Mingro** //COM mi **dai** una rondella //COM ‘'*Hi Mingro. Can you pass me one rondella?*’
%ill: [1] recall; [2] request
*MIC: una **rondella** ?COM ‘*one rondella?’*
%ill: surprise question
*KAT: **sì** //COM ‘*yes*’
%ill: positive assertion
*MIC: **una rondella** //COM ‘*one rondella*’
%ill: confirmation

First of all : given that in (3) the surprise of MIC is not objectively motivated because some ‘rondelle’ are available on the table, the input must have been once more a psychological one. How to find what should be the contextual input, i.e. the covert question, for the reaction of surprise? It seems that the surprise depends on the fact that MIC was absent-minded, and that his attention was not directed to the present context; but it could depend also on his malice against KAT, because the general level of the conversation is quite heated. It is impossible to find proof to decide. Secondarily: how to motivate the pursuit of the dialogue, because KAT answers to the surprise question with a kind of strong assertion confirming his own request, but he should have answered with an expression of impatience or worse. Finally, MIC agrees on the content of the request, allowing the conclusion of the short negotiation, but he should have felt offended. So it seems impossible to find any objective mandatory question- input in the situation that could be reasonable for the development of the dialogue.

The reported examples are not exceptions but the normal manner of human spoken communication which is about the context but which has its origin in the speakers’ thoughts and in the affective dynamics among speakers. They are not determined by context and they continue on with subjective actions and reactions. So the utterance’s information structure cannot be reduced to the semantic packaging of “an answer to a question” or “an answer to covert questions suggested

by the context". In this regard also, the so called pragmatic definition of Focus, in terms of the point marking the possible alternative answer with respect to covert or open questions in the context, seems to be vacuous or not useful in explaining the real nature of speech information structure.

In conclusion there is much evidence from corpus data showing how the context question-answer model is far from the reality. In principle the model had to be satisfactory for utterances with assertive illocution, whose input should be a kind of overt or covert question in the input, but given that at least 40% of the illocutionary values of utterances in spontaneous speech are not assertive, it is not clear what could be the covert question in the context, being the adequate input of different speech acts typology. What is the context- question generating an alternative question, or an instruction, or an expression of obviousness? Moreover even the development of a dialogue from overt questions by the speaker to answers by the addressee, that has to be the normal way, does not occur so frequently, because very often the addressee prefers not to answer and behaves in another way on the basis of his subjective motivations, as it has already been shown. So also this canonical situation may depart from the model. But a constant aspect of every utterance derives from its pragmatic nature and from its illocutionary types which very rarely can be connected in an incontrovertible way to an objective/contextual input, and on the contrary ties to an internal affective disposition.

3.2.4. The Contrastive Focus. At this point it must be stressed that real speech must also be studied considering its sound counterpart and especially some prosodic cues like terminal and non-terminal breaks, prosodic forms with illocutionary values, prosodic prominences signaling necessarily focus. In accordance with these premises, it is assumed by the most part of literature that Focus must correlate with a phonetic-prosodic prominence[25]. The taking into account of this prosodic cue causes new contradictions, because it cannot be ignored, too, that there are utterances bearing two prosodic prominences[26]. Thus, the occurrence in the same utterance of a first prominence and a second one, corresponding to semantic Foci, are a phenomenon it becomes necessary to explain.

In reality systematic controls on the corpus carried out in our Laboratory make us sure that not only the *root* PUs performing a Comment are characterized by a prosodic prominence, but that also *prefix* PUs performing Topics, are mandatorily concluded by a perceptual prominence, sometimes more relevant than that in the

[25] The prosodic aspects of prominences marking Focus will be treated later. See studies on prosodic Focus (Avesani & Vayra 2003; Avesani & Vayra 2004; D'Imperio 2001).

[26] Somebody could hypothesize that the semantic character of one of the two prominences within a same utterance should not be new, nor important, in this way excluding its focal value. But in most cases the semantic content shows enough relevance to confirm its semantic focal position.

Comment. This means that the Topic-Comment IP is always performed with a *prefix* PU and a *root* PU, each of them recording a prominence, corresponding to the prosodic nucleus of the PU. See, for instance, examples (11) and (12). In conclusion, every utterance corresponding to a Topic-Comment pattern is characterized by two Foci.

Facing the case of the two Foci utterances, scholars have been, in some sense, obliged to make the hypothesis of a Contrastive Focus (Büring 2003). This has been explained within the model of context question-answer through the hypothesis of a double question which should motivate the double Focus (*who stole what?*)[27].

It is obvious that if the finding of a mandatory question input in the context to explain a Focus in the answer hardly appears acceptable, the hypothesis that the context questions had to be double to also explain a Contrastive Focus seems even less so. It must be considered moreover that corpus data records about 10% of non-simple topicalisation phenomena, i.e. the IP of a lot of utterances is not compounded of a Topic-Comment pattern, but of a Topic-Topic-Comment, or a Topic-Topic-Topic- Comment, or of a List of Topics and a Comment[28]. In this case each of the *prefix* PUs, performing the respective Topic, bears its own prosodic prominence, marking a Focus. Thus, according to the question-answer model there has to be a new Contrastive Focus every time there is a Topic, and by consequence a multi-multi covert questions input has to be found in the context for justifying that result[29].

For instance, how to formulate for (22) a triple covert questions input, implying a covert question for the first Contrastive Focus in Topic, one for the second Contrastive Focus in the second Topic, and finally one for the Focus in the assertive Comment? More or less the triple covert question suggested by the context has to be: how many are they? what have they brought? are they right?

(22) *MAA: la **maggior parte** /TOP […] quelli che hanno portato **Pinocchio** /TOP va proprio bene quello **che hanno** //COM *'the most part, those who brought Pinocchio, it is all right what they have'*
%ill: assertion [ipubcv02]

[27] It must be noticed that hypothesis of Contrastive Focus indeed does not assume that there is a Focus in the Topic but only that there are utterances with two Foci, one of whom is considered "contrastive".
[28] In this case the expression behaving like the second or the third Topic must be in any case and by itself an adequate field of application for the illocutionary force of the Comment. That is, it must be semantically representative, otherwise it cannot be performed such as a Topic.
[29] On the contrary, conformably to our information perspective the speaker can duplicate or triplicate the field of application of the illocutionary force, Topic, with the explication of linguistic details.

We don't see how it could be possible to justify as input such a triple question in the context. We already have advanced doubts about the question-answer model but when the general frame is also extended to a second or third or more covert questions, it seems to be a totally *ad hoc* solution.

In conclusion: in a lot of influential literature the notion of Focus is strictly semantic and has been considered the central point for the information structure of the utterance. The concept has been traditionally defined according to vague notions of importance and novelty. Starting from the assumption of Common Ground within the model of context question-answer, more recent approaches have proponed the function of Focus as highlighting a semantic alternative in the answer and have hypothesized the existence of Contrastive Foci to explain the occurrence of utterances with two Foci. We have been arguing against this perspective both theoretically and on the basis of corpus data evidence.

3.3 The LAcT definition of Focus

3.3.1. In the LAcT perspective the importance of the concept of Focus is strongly rescaled because the information structure is not conceived as a semantic entity with a propositional size/form, whose Focus has to be the center. Information patterning does not depend on it, but on the pragmatic accomplishment of an illocution by the Comment, and on the pattern of Topic-Comment performing a passage of the information from a semantic representative domain to a necessarily new domain. The overall structure is not semantic but is still informative, because the relation of Topic with Comment is founded on a condition of pragmatic aboutness and leads to an utterance. The definition of which, too, is pragmatic and not semantic, owing to its correspondence to a speech act and not to a proposition.

Focus remains a semantic concept in LAcT too, but its domain spreads only until the boundary of a textual IU of Comment or until one of Topic. Expressions are conceived to develop an information function of Comment or Topic in the performance of the illocutionary act. Simultaneously the same expressions, produced with an information function within the illocutionary act, are performed such as syntactic configurations and semantic islands within the locutionary act.

Inside the locutionary performance, each island is composed according to syntactic and semantic rules. Specifically the semantics of each domain of Topic and each of Comment records different kinds of relations regarding regency, quantification, modification, predication, negation, modality, and Focus. Focus is a high semantic level of composition occurring both in Comment and Topic IUs. So, even if Focus is still a semantic notion, its domain is related not to an entire

utterance, or presumed proposition, but to the semantics of one Topic or Comment domain, which often copes only with syntactic phrasal constituents[30].

Thus a general semantic definition of Focus in LAcT:

> A Focus signals the apex of a semantic domain which develops a Topic or a Comment information function.

The semantics of the domain behaving like a Topic or a Comment is conditioned by the information function that the expression is developing: in the case of Topic that of a field of application of an illocutionary force (T-Focus) and in the case of Comment that of the expression of an illocutionary force (C-Focus). We have already seen that the general semantic condition of a Topic domain is being representative, so that the representation of a pragmatic prominence is allowed. Corpus data shows that 75% of the linguistic content corresponds to Noun phrases and Prepositional phrases, so it is possible to also foresee its most common kind of denotation (individual entities, times, places, modality) and imagine that T-Focus will be the apex of these kinds of domains. As regards the Comment domain there is no effective restriction, beyond the limit of morpheme, in order to develop the illocutionary function. Anyway, it is more appropriate saying that there are more than a condition's semantic preferences accomplishing specific illocutions. For instance all languages have developed idiosyncratic formulas to express ritual illocutions (greetings, regards, thanks, excuses), and it is more frequent that Verb phrases occur to accomplish the communicative actions in Comment than to express their field of application in Topic. So C-Focus often corresponds directly to a formula or to the apex of a semantic domain denoting an event.

Then the Foci of these two different textual IUs are apexes of semantic domains which not only develop different functions but systematically diverge also for their semantic content and their respective lexical and morpho-syntactic composition. Therefore we claim that there are *Topic Focus* (T-Focus) and *Comment Focus* (C-Focus)[31].

Generally speaking, T-Focus has a semantic identification function within a non action domain and C-Focus has a semantic specification function within an action domain.

---

[30] Evidently it may correspond to whatever kind of constituent (even an entire sentence behaving as an IU).

[31] The lack of the notion of Topic Focus in the tradition is a consequence of the lack of the illocutionary nature of the utterance and of the distinction between Focus and Comment, even if a large part of literature considers focalized the expressions with a Topic function.

3.3.2. According to corpus data implying the consideration of sound counterpart, it seems useful to remember that a necessary feature of Focus is that it is marked by a prosodic prominence through different parameters. The most important are:

a) pitch with a perceptually relevant F0 movement (rising-falling, or rising) or a strong modulation movement;
b) duration with the lengthening of the syllables (plus a high intensity value).

In all cases the seat of the prominence is the nucleus of the *prefix* PU or the *root* PU involved.

There is not the space in this work to deal with the phonetic and prosodic details of prominence, but it must be said that not all kinds of prominences performed in an utterance are the seat of a Focus. Actually, different types of prominences are commonly realized in speech, but their values are quite different from those of Focus. They can regard for instance the F0 movements necessarily marking the end of a phonetic group (no more than 7 syllables, Martin Ph. 2009). Also the lexical focalisation, which is due to a semantic intensification on a single word, is realized with a prominence lacking a functional value, and the process of focalisation, which is induced by Focus sensitive particles like 'even', 'also', 'only', 'no', 'not', makes the subsequent word focalised. Otherwise, some Dialogic IUs with high activation, like Incipit, Conative, and Dialogue Connector are performed through PUs with prosodic prominence, or sometimes the non terminal break inside a long scanned IU can be accompanied by a prominence. And finally there can be phenomena of rhythm and "lingua specific melody" which can produce some prosodic modulations.

Quite simply, whatever utterance that's not too short records some prominences, which can be recovered both manually and automatically,[32] but not all of these are marks of Focus. Actually, prominences signaling T-Focus and C-Focus are specific, corresponding to the prosodic Nucleus of a *prefix* PU and a *root* PU, and only these are relevant for our perception in order to identify a field of application or to specify an illocutionary type.

3.3.3. For what regards T-Focus, following its function it must be the apex of a domain adequate to identify the field of application of the illocutionary force. Coming back to (7), in an utterance with a total question force, but with two Topics, each Topic must identify a field of application for the question in Comment and each of them should function as a Topic by itself.

[32] The measures and features of prominences are objects of important research. See for the detection of Focus prominence (Tamburini 2005; Gagliardi 2009; Ph. Martin 2010).

(7) *CEC: di **là** /TOP gli **acidi** /TOP tutto **pronto** ?COM '*there, (for what concerns) the acids, everything ready?*'
%ill: total question [ifamdl17]

The right part of the *prefix* PU is the seat of its nucleus with a prosodic prominence, and the majority of times it is performed with a rising or a rising-falling movement. This position copes with the last semantic word of each Topic. So the adverb 'là' (*there*) and the noun 'acidi' (*acids*) can be considered the respective semantic Focus marked by the prosodic prominence.

It can also happen that if the word involved is a Noun it may be preceded by a quality Adjective, rarely in Italian, or by a grammatical modifier (possessive, numeral, indefinite), or it can be a fixed expression; in the latter case the prominence can include all the modified expressions like in (22) where the entire group 'maggior parte' (*most part*) copes with the prominence.

(22) *MAA: la **maggior parte** /TOP […] quelli che hanno portato **Pinocchio** /TOP va proprio bene quello **che hanno** //COM '*the most part, those who brought Pinocchio, it is all right what they have'*
%ill: assertion [ipubcv02]

Very often the expression, functioning as Topic[33], is from a syntactic point of view a well formed phrase (Noun, prepositional, adverbial, adjectival), whose last word is also the head of the phrase[34]. But it can happen that there is not this coincidence like in the second Topic of (22), where the proper name 'Pinocchio' is the last word but it is not the head of the noun phrase. It should have been in doubt regarding the semantic or syntactic condition for being the Focus of a Topic domain, but corpus examples allow us to verify that it is always the final seat that correlates with the role of Focus in spite of the syntactic head position.

The fact that the last word in the Topic IU corresponds to the semantic apex of the domain, seems to suggest that the relevant feature is that romance languages "build on the right"[35]. We have the habit of expecting the end of something in recognizing it as a whole and in speech the signal of ending or starting is given primarily by prosody. As a result, the last semantic word of the Topic marked by a prosodic prominence is recognized by the hearer as the expression closing the

[33] It must be remembered that there is no syntactic relation between the linguistic filling in both Topics to each other and to that in Comment, so that they can be defined as anacholuta and from a semantic point of view are islands.
[34] The linguistic material of one IU is short in the majority of cases, corresponding to a few words composing a phrase, so often the last word and the head word of the phrase coincide.
[35] We have not enough research to assume that it is the deepness in the syntactic structure that assigns the Focus. There are many examples which seem to contradict this hypothesis.

domain and identifying it as the semantic entity to be considered in its whole the application field of the illocution i.e. the Topic.

In conclusion, T-Focus occurs generally on the last semantic word of the Topic IU, concluding and identifying the semantic entity *allowing the representativeness* of the field of Comment illocutionary force and ensuring the semantic recoverability of the entire entity.

3.3.4. On the contrary C-Focus has no fixed seat, even if it too occurs very often on a semantic word in the right side of a Comment IU or only on the last word. It depends on the fact that the C-Focus is also marked by the nucleus of the *root* PU, but this can occur in different seats within the PU[36]. Below are some examples with different illocutions where the C-Focus doesn't occur on the last word of the IU.

(23) *PAO: il resto /TOP non voglio **sapere** che cosa pensano //COM '*for the rest, I don't want to know what they think'*
%ill: refusal [ipubcv01]

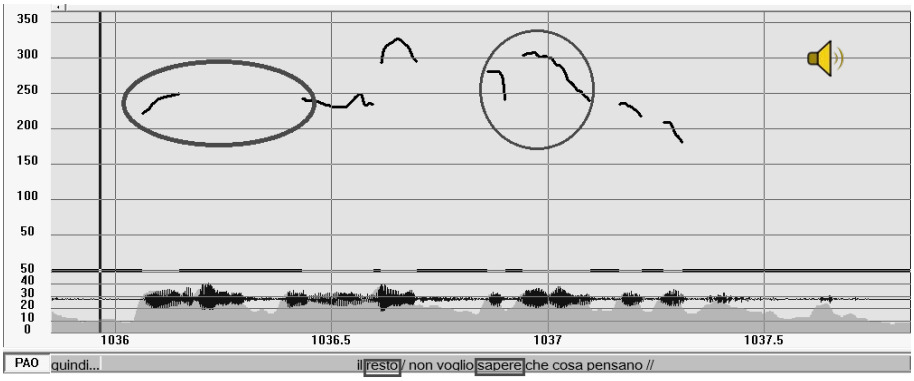


Figure 5.

(16) *LIA: che si doveva fa' **perdonare** /TOP non l' ho **mai** voluto sapere //COM '*that she had to be forgiven, (I) did not want to know it*'
%ill: disagreement [ifamcv01]

[36] Anyway it is not to be confused with the fact that the nucleus of a *root* PU, in accordance with its illocutionary value, can be preceded by a prosodic part of *preparation* or followed by a part of *tail*. The seat of Focus remains in the nucleus.

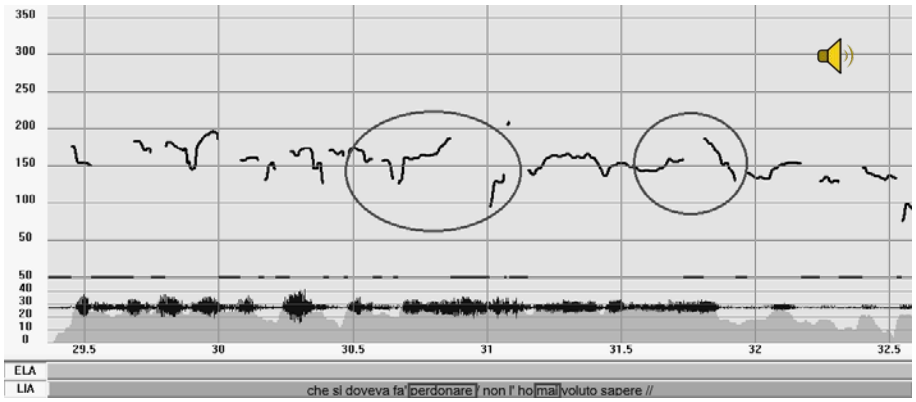


Figure 6.

(24) *VAL: perché io sono stata nominata /SCA **prima** di' trenta di' giugno //COM
'*cause I have been appointed, before the thirtieth of June'*
%ill: answer [ifamvc18]

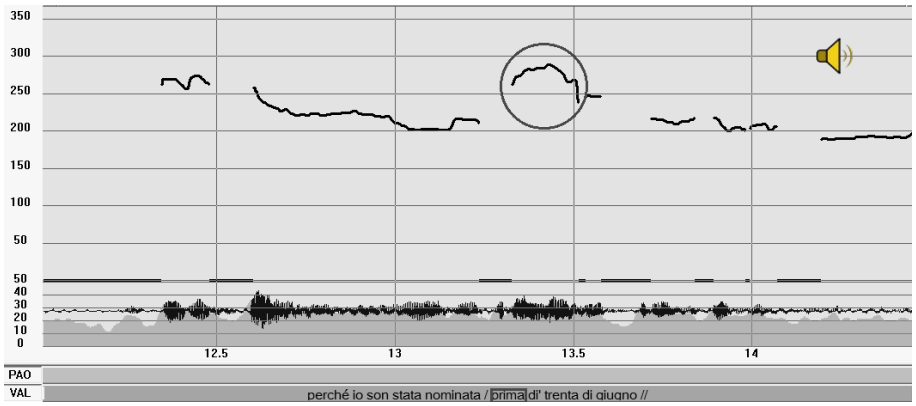


Figure 7.

Anyway for what it has been possible to verify, C-Focus, coping with the prosodic nucleus of the *root* PU, represents the phonetic part necessary to express and specify the illocutionary type of the Comment. It means that the recoverability of the illocutionary type is assured if the only sound of the prosodic nucleus within the *root* PU is conserved[37]. Below are some examples where the listening of the bare nucleus

[37] Evidently in experimental research cutting the rest of the sound of the *root* PU and conserving only the nucleus, if an expert can still recognize the illocutionary value accomplished, the linguistic interpretation of the nucleus is unsatisfactory.

of the *root* PU allows the recognition of the illocutionary value. See (7), where the prosodic shape of a total question is clearly recognizable from the last two syllables.

(7) *CEC: di **là** /TOP gli **acidi** /TOP tutto **pronto** ?COM '*there, (for what concerns) the acids, everything ready?*'
%ill: total question [ifamdl17]

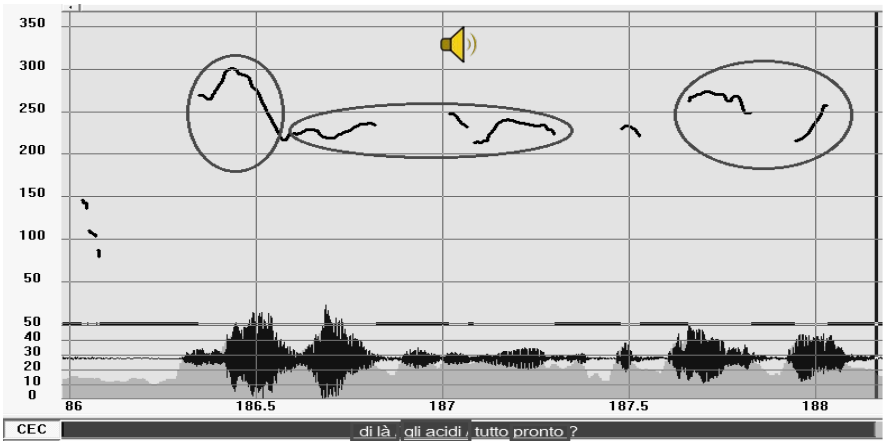


Figure 8.

In (25) a partial question is performed.

(25) *PRO: l'unit linked /TOP **praticamente** /TOP che **cos'è** ?COM '*the linked unit, actually, what is it?*'
%ill: partial question [ipubdl04]

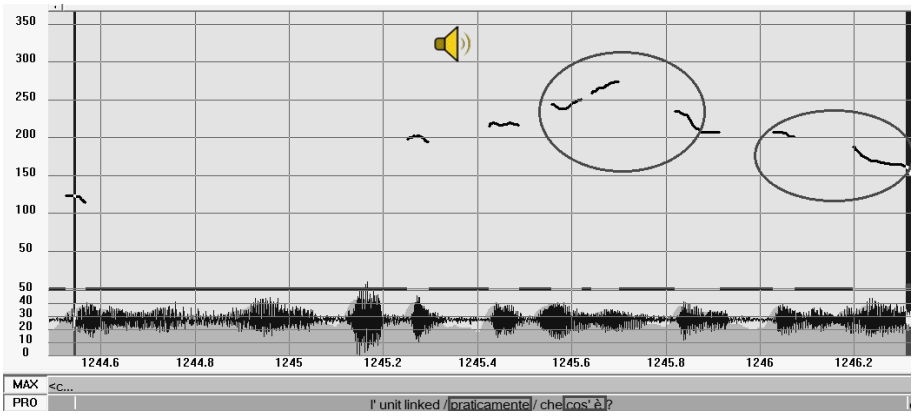


Figure 9.

(26) is an example of the expressive illocution of Contrast with a high jump on *you*.

(26) *PAO: che tu me l'avevi detto **te** /COM i'cream caramel //APC '*cause it was you that said it to me, the cream caramel'*
%ill: assertion of contrast [ifamdl12]

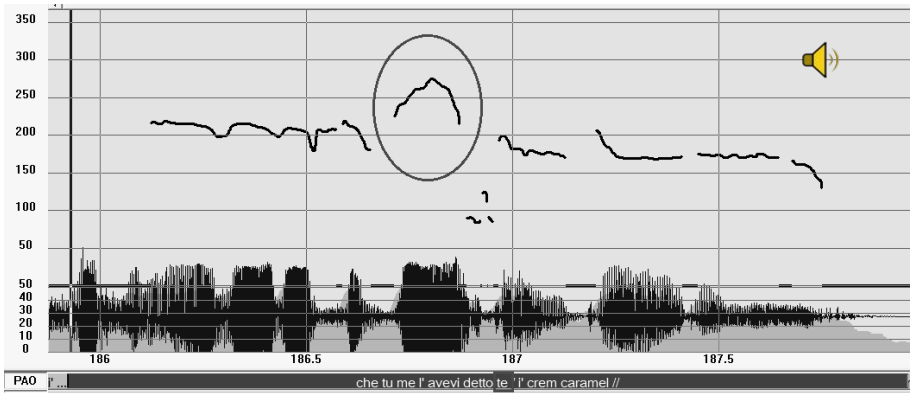


Figure 10.

The correlation between the nucleus of the *root* PU and the seat of Focus can produce peculiar situations like in the case for instance of a partial question, or an expressive illocution. Actually, the prosodic forms of these *root* PUs foresees that the nucleus is spread on an entire group of words and in this case Focus can also cope with all the expressions employed. Let us see (27), (28) and (29).

(27) *EMA: questo periodo /TOP **quanto dura** ?COM '*this period, how long does it take?*'
%ill: partial question [ipubdl05]

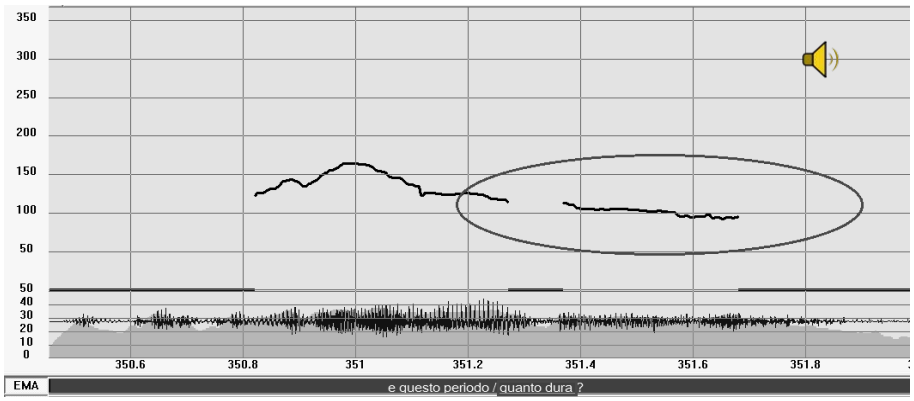


Figure 11.

(28) *FRA: come si dice / i'tiramisù ai'nescaffé /TOP dev' **essere una cosa** …COM
'*you mean, the tiramisù made with nescaffè, it must be…'*
%ill: expression of obviousness [ifamdl12]

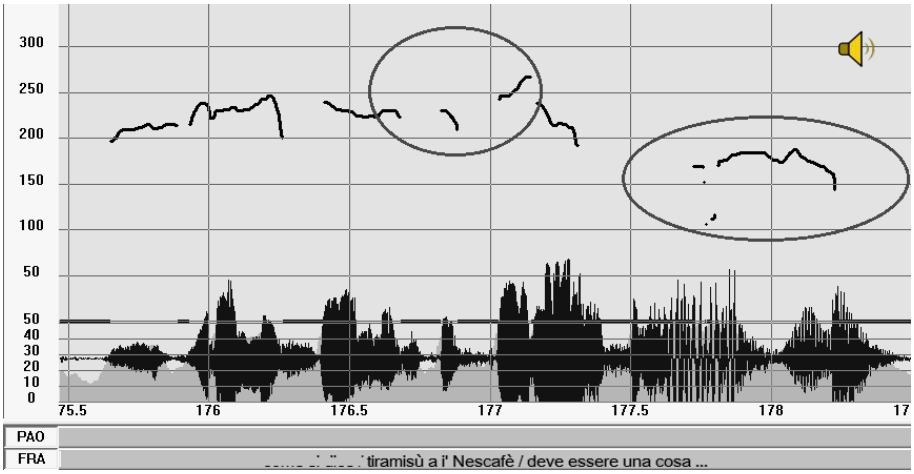


Figure 12.

(29) *ZIA: **questa voce** //COM '*this voice'*
%ill: evocation [LABcorpus]

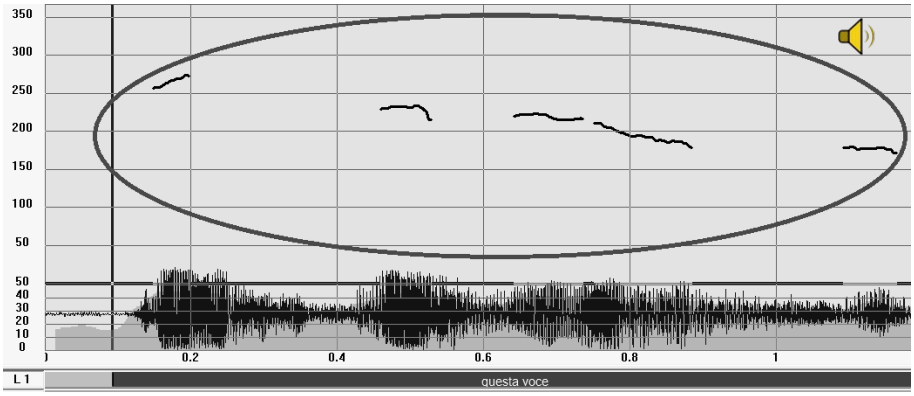


Figure 13.

Evidently what is relevant to perform with a C-Focus more than the recoverability of a semantic domain, like in the case of T-Focus, is the sense of an expression through which a specific act is accomplished. Then the goal of C- Focus emerges for supporting the word (s) and bettering their sense with which a specific illocution can be recognized, so doing, it prompts the addressee's attention to the latter.

C-Focus *marks the expression allowing us to specify what type of illocution is performed* within the semantic domain, dedicated in its whole to the accomplishment of the illocutionary force.

But scholars know that the prosodic prominence marking the nucleus, coping with the C-Focus, can sometimes be of little relevance and specifically it can be less strong than that of T-Focus, if an IU of this type occurs in the IP of the utterance. The difference can also be easily appreciated from our examples.

This is not so strange, all things considered, because it is understandable that the perceptual prominence of T-Focus is more relevant than that of the C-Focus within a Topic-Comment pattern. It depends on the fact that what is necessary and mostly relevant is that the *root* PU of Comment must clearly manifest a specific illocution. So this task is accomplished more by the form of the *root* PU than by the scale of the prosodic prominence and by its apex, while in the Topic the only way to signal the Focus is through the relevance of its prominence.

A Focus occurs on the word culminating the illocution of the speaker's intention: for instance in (16) the Topic introduces the field of forgive and the Comment, with a disagreement illocution, and finds its Focus on the negative adverb never which is in the middle of the expressions. In this way the final sense is the total disagreement about a possible forgiving of something.

(16) *LIA: che si doveva fa' **perdonare** /TOP non l' ho **mai** voluto sapere //COM
'*that she had to be forgiven, (I) did not want to know it*'
%ill: disagreement [ifamcv01]

In conclusion the IP of the utterance has a pragmatic nature and its origin is in the accomplishment of an illocutionary force by the Comment. IP does not correspond to a semantic structure whose center is the Focus. IP does not depend on Context, and also Focus does not. Focus corresponds to a semantic level of composition within the domain of a Topic and a Comment IUs, and while T-Focus develops the specific function of allowing the representativeness of the field of illocutionary force, in its turn the function of C-Focus is allowing the specification of the illocutionary type. Both are mandatorily signalized by the nuclear prominence of their respectively prefix PU and root PU.

References

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

Avesani, C. & Vayra, M. 2003. *Broad, narrow and contrastive focus in Florentine Italian*. In M.J. Solè, D. Recasens & J. Romero (eds), *Proceedings of the 15th International Congress of Phonetic Sciences* vol 2. Barcelona: Causal productions Pty, 1803-06.

Avesani, C. & Vayra, M. 2004. Focus ristretto e focus contrastivo in italiano. In A. Leoni, F. Cutugno, M. Pettorino & R. Savy (eds), *Il Parlato italiano. Atti del Convegno Nazionale.* Napoli: D'Auria Editore, 1-20.

Bally, C. 1932. *Linguistique Générale et Linguistique Française*. Berne: Francke Verlag.

Berruto, G. 1985. Dislocazioni a sinistra e Grammatica dell'italiano parlato. In A. Franchi De Bellis & L.M. Savoia (eds), *Sintassi e morfologia della lingua italiana d'uso. Teorie e applicazioni descrittive*. Roma: Bulzoni, 59-82.

Büring, D. 2003. On D-trees, beans and B-accents. *Linguistics and Philosophy* 26: 511-545.

Chafe, W. 1970. *Meaning and the structure of language.* Chicago/London.

Chafe, W. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C.N. Li (ed.), *Subject and Topic*. New York: Academic Press, 25-55.

Chomsky, N. 1971. Deep Structure, Surface Structure and Semantic Interpretation. In D. Steimberg & L. Jacobovits (eds), *Semantics: an Interdisciplinary Reader.* Cambridge: Cambridge University Press, 183-216.

Cresti, E. 1987. L'articolazione dell'informazione nel parlato. In *Gli Italiani Parlati: Sondaggi sopra la lingua d'oggi.* Firenze: Accademia della Crusca, 27-90.

Cresti, E. 1994. Information and intonational patterning in Italian. In B. Ferguson, H. Gezundhajt & P. Martin (eds), *Accent, Intonation, et Modéles Phonologiques.* Toronto: Editions Mélodie, 99-140.

Cresti, E. 2000. *Corpus di Italiano Parlato*, voll. I-II, CD-ROM. Firenze: Accademia della Crusca.

Cresti, E. 2006. Some comparisons between UBLI and C-ORAL-ROM. In Y. Kawaguchi, S. Zaima & T. Takagaki (eds), S*poken Language Corpus and Linguistics Informatics.* Amsterdam: Benjamins, 125-152.

Cresti, E. (forthcoming). L'unité de suffixe: identification et interprétation des unités de la langue parlé. In S. Caddéo, M.N. Roubaud, M. Rouquier & F. Sabio (eds), *Penser les langues avec Claire Blanche-Benveniste.* Presses Universitaires de Provence.

Cresti, E. & Firenzuoli, V. 1999. Illocution and intonational contours in Italian. *Revue Française de Linguistique Appliquée* IV (2): 77-98.

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

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

Cresti, E., Moneglia, M. & Tucci, I. 2011. Annotation de l'entretien d'Anita Musso selon la Théorie de la langue en acte. In F. Lefeuvre & E. Moline (eds), *Langue Française* n. 170, 95-110.

C-ORAL-ROM. http://lablita.dit.unifi.it/coralrom/

D'Imperio, M. 2001. Focus and tonal structure in Neapolitan Italian. In *Speech communication* vol. 33, issue 4. Amsterdam: Elsevier Science Publishers, 339-356.

Firenzuoli, V. 2003. Le forme intonative di valore illocutivo nell'italiano parlato. Analisi Sperimentale di un Corpus di Parlato Spontaneo (LABLITA). PhD diss., University of Florence.

von der Gabelentz, G. 1891. *Die Sprachwissenschaft, ihre Aufgaben, Methoden und bisherigen Ergebnisse.* Tübingen.

Gabriel, Ch., Lleò, C. (eds) 2011. Intonation Phrasing in Romance and Germanic, Amsterdam: Benjamins.

Gagliardi, G. 2009. Correlati fonetico-acustici dell'informatività. Thesis. Università di Bologna.

Grice, H. 1975. Logic and Conversation. In P. Cole & G. Morgan (eds), *Speech Acts. Syntax and semantics* vol. 3. New York: Academic Press, 41-58.

Gundel, J.K. 1977. *The Role of Topic and Comment in Linguistic Theory.* Indiana University Linguistics Club.

Halliday, M.A.K. 1976a. *Intonation and grammar in British English.* The Hague: Mouton.

Halliday, M.A.K. 1976b. *System and Function in Language: Selected Papers*. London: Oxford University Press.

't Hart, J., Collier, R. & Cohen, A. 1990. *A Perceptual Study on Intonation. An Experimental Approach to Speech Melody*. Cambridge: Cambridge University Press.

Hockett, C. F. 1958. *A Course in Modern Linguistics*. New York: The Macmillan Company.

Jackendoff, R. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge Mass: MIT Press.

Jacobs, J. 1984. Funktionale Satzperspektive und Illokutionsemantik. *Linguistische Berichte* 91: 25-58.

Lambrecht, K. 1994. *Information structure and sentence form*. Cambridge: Cambridge University Press.

Krifka, M. 2001. For a structured account of questions and answers. In C. Féry & W. Sternfeld (eds), *Audiatur vos sapientiaer. A festschrift for Achim von Stechow*. Berlin Akademie: Verlag, 287-319.

Krifka, M. 2006. Basic Notions of Information Structure. In C. Féry & W. Sternfeld (eds), *Audiatur vox sapientiae: A festschrift for Achim von Stechow*. Berlin: Akademie Verlag, 287-319

van Kuppervelt, J. 1994. Topic and comment. In R.E. Asher (ed.), *The Encyclopedia of Language and Linguistics.* Pergamon Press, 4629-4633.

LABLITA. http://lablita.dit.unifi.it/

Li, C. (ed.) 1976. *Subject and Topic*. New York: Academic Press.

Martin, Ph. 2009. *Intonation du Français.* Paris: Armand Colin.

Martin, Ph. 2010. Prominence detection without syllabic segmentation. In *Speech Prosody, Satellite work-shop on prosodic prominence*. http://www2.unine.ch/webdav/site/speechprosody-prominence/shared/proceedings/Prominence_2010_MARTIN.pdf

McWhinney, B. 1994. *The CHILDES Project. Tools for analyzing Talk*. Hilsdale, N.J: Lawrence Erlbaum Associates.

Moneglia, M. 2011. Spoken corpora and pragmatics. In H. Mello & S. Gries (eds), *Brasilian Journal of applied linguistics, Revista brasileira de linguìstica aplicada* vol. 11, n. 2, 479-519.

Moneglia, M. & Cresti, E. 1997. Intonazione e criteri di trascrizione del parlato. In U. Bortolini & E. Pizzuto (eds), *Il progetto CHILDES Italia* vol. II. Pisa: Il Cerro, 59-90.

Paul, H. 1880. *Prinzipien des Sprachtgeschicte*. Leipzig.

Reinhardt, T. 1982. Pragmatics and linguistics: an analysis of Sentence Topics. *Philosophica* 27: 53-94.

Rooth, M. 1992. A theory of focus interpretation. *Natural Language Semantics* 1: 75-116.

Tamburini, F. 2005. Identificazione automatica della prominenza frasale nel linguaggio parlato. In P. Cosi (ed.), *Misura di Parametri. 1° Convegno nazionale AISV*. Brescia: EDK Editore, 725-754.

Searle, J. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Searle, J. 1975. Indirect speech acts. In P. Cole & J.L. Morgan (eds), *Syntax and Semantics 3: Speech Acts.* New York: Academic Press, 59-82.

Sornicola, R. & Svoboda, A. 1989. *Il campo di tensione*. Napoli: Liguori.

Signorini, S. 2005. Topic e Soggetto in corpora di italiano parlato spontaneo PhD diss., Università di Firenze.

Stalnaker, R. 1974. Pragmatic presuppositions. In K. Milton, M. Munitz & P. Unger, *Semantics and Philosophy*. New York: New York University Press, 197-214.

Weil, H. 1844. De l'ordre des mots dans les langues anciennes comparées aux langues modernes. In *The order of words in the ancient languages compared with that of the modern languages translation*. Amsterdam: Benjamins.

ILLOCUTION AND MODALITY IN SPOKEN ITALIAN: PERFORMING A SPEECH ACT THROUGH WORDS AND JUDGING THEIR SEMANTIC CONTENT

A CORPUS-BASED ANALYSIS

Ida Tucci
LABLITA - University of Florence

1. Premise

Despite the long tradition of studies on Modality, this notion is not clearly distinguished from Illocutionary force, which many authors still consider in terms of "modality of the utterance" (assertive, deontic, evaluative, etc.) making the semantic value of the utterance collapse on his pragmatic result. This problem is sensible for spontaneous speech processing, where on one side modal lexical indexes are frequent and, on the other, speech act analysis is crucial for parsing the speech flow.

The paper discusses the results of a research in which lexical modal indexes have been retrieved in the C-ORAL-ROM Italian corpus (Cresti & Moneglia 2005) and then annotated according to their modal values (Alethic, Epistemic and Deontic) and illocutive classes (Representatives, Directives, Expressives, Refusals and Rytes). The distribution of modal indexes, face to the informational and pragmatic structure of spoken Italian, clearly shows that Modality and Illocution are two independent levels of the utterance, so they has to be considered different notions.

Looking at the actual data, and in accordance with the *Lingua in Atto Theory* (Cresti 2000), we will show that Illocutionary force and Modality are necessarily independent notions for two main reasons. The first one, presented in 4., is formal. Each utterance has, by definition, one and only one Illocutionary force, i.e. accomplishes one illocutionary act. On the contrary, more than one modal values can coexist in the utterance. In other terms, while Illocutionary force is a property of the utterance, Modality is a property of the elements that compound its structure

Mello H., Panunzi A., Raso T. (eds), *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation*

(information units). From a more qualitative point of view we will show in 5. that, although modal indexes may contribute to the illocutionary interpretation of an utterance, there is no direct correspondence between modal values and Illocutionary forces. More specifically, we will show that each Illocutionary type recorded in the corpus can express all possible modal values and vice versa.

2. Introduction

In the history of philosophy of language and linguistics, many definitions of 'modality' have been proposed. The most classic characterization of this notion dates back to Bally (Bally 1932; Bally 1942), who defines modality as «la forme linguistique d'un jugement intellectual, d'un jugement affectif ou d'une volonté qu'un sujet pensant énonce à propos d'une perception ou d'une représentation de son esprit», i.e. a *Modus* on a *Dictum*, or "the speaker's cognitive, emotive, or volitive attitude towards a state of affairs", his "commitment or detachment", his "envisaging several possible courses of events" or his "considering of things being otherwise" (cf Kiefer 1994: 2516a).

This basic idea has been reported, with many other characters, also in Lyons (1977: 452): «[modality represents] the speaker's opinion or attitude towards the proposition that the sentence expresses or the situation that the proposition describes […]», and in Bybee & Fleischman's views (1995: 2): «When the proposition of an utterance in the most neutral semantic status, i.e. factual or declarative, is subject to further addition or overlay of meaning, this extension represents modality». The concept of *further addition* or *overlay of meaning* in a "neutral" utterance hints towards a lexicalization of modal meanings in languages, that modifies non-modalized entities.

The identification of a subjective/evaluative *attitude* in the actual language performance is, however, puzzling. Palmer (1986; 2001), for instance, presents a general survey of modality as a typological category. He draws attention to the subjective nature of modality, defining it as "the grammaticalization of speakers' (subjective) attitudes and opinions" (Palmer 1986: 16; cf Greene 2007).

The term 'modality' refers to concepts such as *possibility*, *necessity*, *belief* and *volition*. The linguistic expressions conveying these semantic values (such as modal verbs, belief verbs, judgment adverbs, etc.) allow the speaker to qualify what he is saying as "possible", "necessary", "in agreement with his beliefs or wills", etc.

From this point of view, both in ordinary language and in modal logic, Modality has been considered the highest level of the semantic organization of a proposition (or a sentence) and also a way to express in it the "subjectivity" of the speaker (Bally 1932; Bally 1942; Hare 1961; Huges & Cresswell 1968; Lyons 1977; Palmer 1990; Bréal 1987; Traugott 1989; Hengeveld 1988).

A body of recent influential literature attempts to restrict Modality to a contrast between *factual* and *non factual*, or *realis* and *irrealis* (Mithun 1995: 173). Other definitions are strictly grammatical (Huddleston 1984: 164; James 1986; Vogeleer et al. 1999; Diewald 2001: 25). In this view 'modality' is just another name for 'mood'; that is a verbal category which expresses the degree of reality assigned to a sentence (indicative is certain, subjunctive and conditional are uncertain, imperative express orders, etc.).

For what regards spoken language, the counterpart of a proposition (or of a sentence) is necessarily a pragmatic entity, i.e. "the utterance" (cf Cresti & Moneglia 2006 and references therein). Therefore, the relation between modal and pragmatic notions is a crucial field for understanding the nature and the role of the words in a speech context. More specifically, is to be put into consideration if in the utterance Modality regards at the same time the text of an utterance and his pragmatic aim, since both modality and illocution communicate a speaker's 'attitude'.

What is generally claimed is that a speech act content typically includes an indefinite range of modal propositions, which can be *asserted*, *judged*, *interrogated*, *requested*, etc. (Kärkkäinen 1987: 151; Graffi 1994: 100; Schneider 1999: 13). But this view explicitly mixes pragmatic notions (assertive, imperative, interrogative, etc.) with semantic concepts (possibility, necessity, belief, volition, etc.), and Modality and illocutionary force may overlap in the language analysis for what regard their formal indexes. For instance, expressions of volitions can lead to assign an imperative value to the utterance, and, in parallel, an imperative utterance could receive a modal association with expression of volition.

In the background framework of this research the *illocutionary force* of an utterance concerns the attitude of the speaker towards the interlocutor (cf Cresti 2000; Cresti 2003), so the *Modus towards the Partner*, and leads to the performance of pragmatic entities; i.e. linguistic actions. On the contrary, *Modality* consists in the evaluation of the speaker towards his own verbalization, i.e. the *Modus on the Dictum*. In other words, Modality corresponds to the cognitive process of evaluating a proposition, while Illocutionary force give rise to the speech act performance. Therefore, in our view Modality belongs to semantics and not to pragmatics.

3. The definition of modal values

In order to study Modality in the language performance and to retrieve the actual use of modal indexes in ordinary speech, we selected, according with the tradition, the

three main modal values that can be expressed in a proposition: *alethic*, *epistemic* and *deontic* modality (cf Cresti 2002)[1]. The following are their definitions.

Alethic modality: the point of view of the speaker refers to the necessity or possibility of the truth of the propositions - that is to propositions that can be verified in the actual world or in possible worlds, by virtue of logical, factual or perceptual judgments (Lyons 1977; Perkins 1983; Kiefer 1994). The overall definition of the Alethic modality also includes *dynamic modality* (Palmer 1990; Huddleston & Pullum 2002), which expresses ability/disposition of a subject to do something in a possible world.

(a) A leopard *must* be spotted.
(b) A swan *can* be black.
(c) An athlete *is able to* run faster than you!

Epistemic modality: the point of view of the speaker refers to the possibility or necessity of a proposition to be verified in those possible worlds that are specifically related to the speaker's beliefs, opinions or attitudes (*evaluative modality*) (Lyons 1977; Venier 1991; Hoye 1997; Papafragou 2001).

(d) Mario *may leave* tomorrow. (he told me something like that)
(e) Mario *has to be gone*. (I don't see him yet)
(f) Mario is depressed, *I believe*.
(g) *Unfortunately* Mario lost his keys.

Deontic modality: the point of view of the speaker, intended as morally responsible agent, refers to the duty, obligation, permission, wishes that are expressed into a proposition (von Wright 1951; Conte A.G. 1977; Conte M.E. 1995; Pottier 2000). Deontic values extend also to "duties" that are social or moral "obligations" in relation to an axiological manifestation of attitudes (Hare 1961; Galvan 1991).

(h) We *must* finish this work in three days.
(i) I *want* to see him before leaving.
(l) You *can't* lie again to me!

[1] There are several proposed modal typologies which vary from this tripartite option. For instance, Mindt (1995: 45) distinguishes 17 modal meanings: (i) possibility/high probability, (ii) certainty/prediction, (iii) ability, (iv) hypothetical event/result, (v) habit, (vi) inference/ deduction, (vii) obligation, (viii) advisability/desirability, (ix) volition/intention, (x) intention, (xi) politeness/down toning, (xii) consent, (xiii) state in the past, (xiv) permission, (xv) courage, (xvi) regulation/prescription, (xvii) disrespect/insolence.

4. The identification of modal values in a speech corpus

The reference corpus for this research is the C-ORAL-ROM-Italia corpus (35.628[2] utterances, cf Cresti & Moneglia 2005). All utterances in this corpus containing an explicit lexical and/or morphological index of Modality has been retrieved and analyzed. According to the tradition, the following items have been identified:

- Modal verbs
- Belief verbs
- Periphrastic and analytic forms
- *To seem*, *to appear*
- Desire and necessity verbs
- Evaluative adjectives in nominal predicates
- Judgment adverbs
- Verbal moods: Indicative future, Conditional

This lexical strategy in the study of Modality corresponds to a practical need of corpus based analysis, as the above indexes can be retrieved on a formal basis. However, this strategy does not entail that in spoken language the speaker can express his *Modus* on the *Dictum* only by mean of lexical or morphological cues.

Other information may express Modality in speech, such as prosodic cues, facial expressions, gestures, etc. Nevertheless, the retrieval of lexical and morph-syntactic indexes entails that, if an explicit modal index occurs in an utterance, than a modalization also occurs and therefore a solid descriptive basis for the analysis of Modality is ensured. In other words, lexical and morph-syntactic indexes are a sufficient (but not a necessary) index of Modality. Each utterance containing at least one of the above indexes has been then studied for what regard its modal value and its informational structure. From a quantitative point of view, lexical modal indexes are very frequent in spoken language: we found 5.152 lexically modalized utterances, corresponding to 14,5% on the total utterances. Given that in the same corpus subordination interests 20% of utterances, coordination 17%, and verbal negation 11%, lexical modalization can be considered a high frequency feature of the spoken language performance (Tucci 2007; Tucci 2009).

The following are examples of utterances bearing modal indexes that have been classified in accordance to the definitions given in the previous paragraph[3].

---

[2] 1.661 utterances of the Man-Machine interactions were excluded from our analysis.

[3] In the C-ORAL-ROM dialogue annotation system three capital letters preceded by an asterisk (*ABC:) correspond to the speaker's label. Double (//) and single (/) slashes respectively refer to *terminal* or *non-terminal* prosodic breaks. The file in the C-ORAL-ROM collection containing each example is reported in square brackets after the translation. Modal indexes are here in italics.

Alethic modality

(1) *CRI: tirandosi dietro la porta / può aver rimbalzato // [itelpv13]
[when he pulled back the door / it can have bounced //]

(2) *PMA: il pubblico ministero *deve* illustrare i fatti // [inatla01]
[a Public Prosecutor / must illustrate the facts //]

Epistemic modality

(3) *PAL: *dovrebbero* essere in sei / a mangiare // [ifamcv04]
[they should be six / eating //]

(4) *GIA: è / *immagino* / un lavoro allucinante // [ifamdl16]
[well / it is / I suppose / an horrible job //]

(5) *GUI: è andato via / *fortunatamente* // [ifammn22]
[he went away / fortunately //]

Deontic modality

(6) *DAN: ci sono militanti di partito / che *possono* averlo e *hanno il diritto* di averlo // [ipubcv01]
[there are party militants/ that can have it and who have the right to have it //]

(7) *ALE: noi *vogliamo* che l'imposizione si riduca // [ifammn22]
[we want the levy to be reduced //]

(8) *ROS: *bisogna* essere ironici / perché *non bisogna* mai prendere troppo sul serio / quello che ci succede // [imedin01]
[we must be ironical / because we must never take too seriously / what happens to us //]

It must be further clarified that, in natural languages, a specific lexical index of modality does not strictly select one and only one modal value. The value of modal indexes may vary according to holistic factors (linguistic context, communicative context, background assumptions etc). For instance, the modal verb 'deve' (*must*) in (2) could in principle be considered as an index of Deontic modality. But this is not the case in the actual context of (2). The speaker, who is a Public Prosecutor, is talking of its duties in the trial ("the Public Prosecutor *must* illustrate facts") and therefore refers to a "propositions that can be verified in the actual world". Hence, in the previous context, the utterance falls within the definition of Alethic and have been classified in accordance with this value. For this reason the modal value in all

utterances bearing modal indexes was determined applying the previous definitions to the full set of available information and no one to one correspondence between Modal indexes and Modal values have been established.

5. Modality and the pragmatic structures of spoken language

In C-ORAL-ROM and in the *Lingua in Atto Theory* (Cresti 2000) here considered as general frames for the study of spoken language corpora, the utterance is "the minimal linguistic entity such that it can be pragmatically interpreted; i.e. the linguistic entity that is 'concluded' and 'autonomous' from a pragmatic point of view" (Cresti & Moneglia 2006: 91). This definition, that goes in the direction of Austin's perspective (Austin 1962), does not imply any necessary correlation between "utterance" and "proposition", but rather highlights the relations between prosody and the accomplishment of speech acts. Prosody marks the speech act boundaries with *terminal breaks* and is strictly necessary to express the Illocutionary force of utterance.

This feature is crucial to study spoken language. No matter if the locutive content copes with a proposition or if it is in "primitive form" (according to the original Austin's terminology): once prosody specifies how this content is related to the world, then it is an independent utterance. Moreover, in this framework, the correlation between prosody and the linguistic structure of the utterance goes beyond of this. An utterance corresponds to an *informational pattern*, which is isomorphic to a prosodic pattern. That means each prosodic envelope of the utterance is assumed to cover a specific functional role. The set of functional roles identified within the *Informational Patterning Theory* is defined in a closed list of types, which is reported here in Table 1[4].

[4] See the following specific studies for properties and functions of the information units: Firenzuoli & Signorini (2003), for the Topic unit, Frosali (2008), for Dialogical units, Tucci

Table 1. Functions of the Information Units

	Type	Informative function	*Tag*
Textual Informational units	*Comment*	Specifies the illocutionary force of the utterance	COM
	Topic	Specifies the application field of the Comment; i.e. the object, state or event the speech act is about	TOP
	Parenthetical	Inserts meta- linguistic evaluations over the text of the utterance	PAR
	Locutive Introducer	Marks the reported speech, exemplifications, listing, etc.	INT
	Appendix	Integrates the text of the Comment or of the Topic, with non essential information	APC/ APT
Dialogical Informational units	*Incipit*	Signals the turn-taking by the speaker	INP
	Phatic	Regulates and controls the communication channel	PHA
	Allocutive	Alerts the interlocutor	ALL
	Conative	Pushes the interlocutor to take part to the exchange	CON
	Expressive	Stimulate the interlocutor to a common point of view on the utterance	EXP

The idea that one an only one prosodic unit (the Comment) plays the informational role to specify the illocutionary force of the utterance is the core notion of this approach. Because of its function the Comment is the only unit that is always necessary (and sufficient) to accomplish a speech act. Therefore an utterance can be simple; i.e made up by one information unit (necessarily a Comment), or compound, i.e. a patterned in different information units (a Comment unit plus others IUs).

In this research, in order to investigate the relation between Modality and the utterance in spoken Italian corpora, the annotation of the information units type has been added to all Modal utterances retrieved in C-ORAL-ROM (Tucci 2007; Tucci 2009). As a consequence of this annotation it has been possible to know in which information unit modal indexes are placed.

An important result of this corpus-based research is that information units show strong regularities and preferences for what regard lexical modalization. In the reference corpus only *Comment, Parenthesis*, *Topic* and *Locutive introducer* contribute to the expression of modality into the utterance. On the contrary, neither *Dialogical units* nor the *Appendix* bear any lexical or morphological cue of Modality. More specifically, modalization occurs by preference in *Comment* and *Parenthesis*, and with a less degree of probability in *Topic* and in *Locutive introducers* (Figure 1).

(2004) and Tucci (2010), for Parenthesis, Firenzuoli (2003) for the typologies of the Comment units.

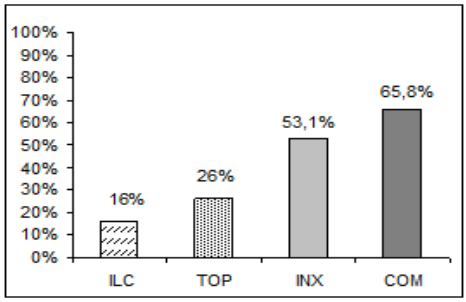


Figure 1. Percentage of information units bearing modal index

It must be noted that on 5.152 utterances bearing a modal index, 3.648 are complex, i.e. compound by more than one *Comment* unit. Moreover, in 2.984 utterances of this set more then one information unit bore a modal index. That means in spontaneous speech many indexes of modality are frequently applied to the utterance and, more specifically, those indexes are distributed over its informational structure. The following are typical item:

(9) *IDA: *in realtà* Basilicata /^{TOP} *dovrebbe significare* la terra dei boschi //^{COM}
[*Actually* "Basilicata /^{TOP} it *should mean* the land of woods //^{COM}] [ifamdl18]

(10) *CLA: *poteva* esse' interpretato così /^{COM} *probabilmente* //^{PAR} [ifammn03]
[It *might have been* interpreted in this way /^{COM} *probably* //^{PAR}]

The distribution of modal indexes across information units has strong theoretical relevance for the study of the relation between Illocutionary force and Modality in spoken language. Indeed, given that in written language and in formal languages Modality is a *property of a proposition*, it might be expected that in spoken language Modality is a *property of the utterance*. But, as the close analysis of the previous example will show, this is not the case.

In (9) the *Topic* bears is an index of Alethic modality, 'in realtà' (*actually*), while in the *Comment* unit the Conditional Mood indicates an Epistemic modality. In (10) an Epistemic index is placed in *Comment* and it is joined to an Alethic modality in *Parenthesis*. Which is the modal value of the above utterances? Do they have Alethic or Epistemic value? This question should not be puzzling in written language and especially in Modal Logic, where Modals are strictly compositional. For instance the following propositional counterparts of (9) and (10) have an Alethic modality. This is caused by the fact that the Epistemic index falls within the scope of the Alethic index:

(9') È *nei fatti* vero che *io credo* che il termine "Basilicata" significhi "terra dei Boschi"
[It is factually true that I believe the term "Basilicata" to mean "land of the woods"]

(10') È fattualmente probabile che nella mia opinione l'interpretazione sia questa
[In my opinion is factually probable that interpretation]

But this is not the case in speech. The above propositions are not possible paraphrases of (9) and (10) which do not have these meanings. In (9) the speaker "adds" a modal Epistemic character to his factual premise, weakening it. The speaker does the reverse in (10), "adding" a factual judgement of probability to his early supposition. Therefore, in the actual interpretation of (9) and (10) the modal indexes do not generate compositionally one modal value, but the scope of each modal index is limited by the information unit boundaries. This is not obviously the case for what regard the Illocutionary force that, by definition, regards the all utterance (*declarative*, in both cases). Finally, looking to the interpretation of modal indexes in patterned utterances, we must conclude that, contrary to Illocutionary force, the Modal value is not a property of the utterance, but rather it is a property of the information unit.

6. Relations between modality and Illocution in the Comment unit

As a consequence of this tagging in our reference corpus, the distribution of modal values over illocutionary values can be explicitly observed. Given that one utterance may have more than one modal value, the distribution only considers utterances bearing Modality in the Comment unit, that who express the Illocutionary force.

To the ends of this paper, is necessary to underline that in spontaneous speech the number of illocutionary types does not correspond just to Assertive, Interrogative and Request, as usually considered in traditional language descriptions. On the contrary, the analysis carried out during the last decade based on our Italian corpora has led to the identification of a larger set of about 90 speech act types in speech (Cresti & Firenzuoli 2001; Firenzuoli 2003). These types have been gathered in five general classes that roughly correspond to the searlian taxonomy. Table 2 below lists the illocutionary types under each class (cf Moneglia 2011).

Table 2. LABLITA Corpus Based Reference Table of Speech Acts Classes and Types

Representa-tives	**Directives**	**Expressives**	**Rites**	**Refusals**
Concluding	Distal recall – not visible object	Exclamation	Thanks	
Make assertion	Distal recall – visible object	Expression of contrast	Greetings	
Answering	Proximal recall	Expression of obviousness	Apologies	
Commentary	Distal deixis	Softening	Welcome	
Strong assertion	Proximal deixis	Expression of surprise	Congratula-tion	
Identification	Presenting (object / event)	Expression of fear	Wishes	
Verification	Introducing (person)	Expression of relief	Compliments	
Claim	Request information	Expression of uncertainly	Declaration of legal value	
Hypothesis / Supposition	Request of action	Expression of doubt	Condemnation	
Explanation	Order	Expression of certainty	Condolences	
Inference	Total question	Expression of wish	Baptism	
Definition	Partial question	Expression of disbelief	Promise	
Narration	Alternative question	Expression of pitty	Bet	
Describing	Request of confirmation	Irony		
Quotation	Reported speech	Regret		
Objection	Announcing	Complaint		
Confirmation	Advising	Imprecation		
Approval	Warning	Insinuation		
Disapproval	Suggestion	Derision		
Agreement	Proposal	Provocation		
Disagreement	Recommend	Reproaching		
	Invite	Hint		
	Prompt	Encouragement		
	Permit	Assuring		
	Authorize	Threatening		
	Prohibition	Giving up		
	Instruction			

Therefore each modal utterance in the reference corpus was classified as an instance of one illocutionary class[5]. From the distribution we can first observe that the set of modal utterances of our corpus records utterances belonging to all illocutionary classes. Therefore, there is no pre-theoretical restriction on the relation between modality and illocutionary classes.

Given this preliminary result we can notice however in Figure 2 that modal utterances are distributed with different percentage in five illocutionary classes. Only a few modal utterances were found for Rites and Refusals, which for this reason will not be considered in the following argument.

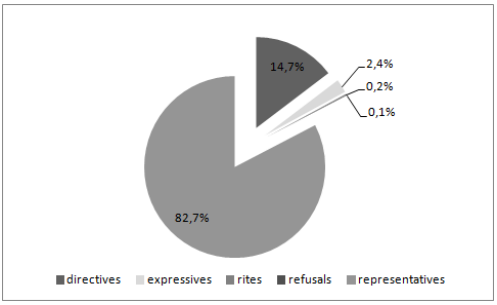
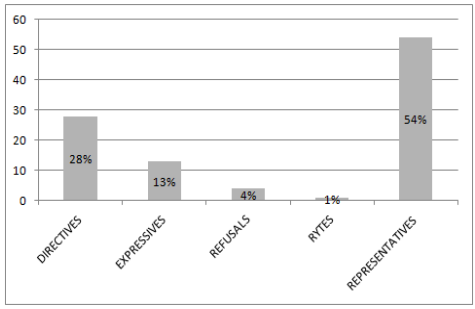


Figure 2. Percentage of Illocutionary classes in the Corpus of Modal Utterances (Tucci 2007)

Firenzuoli (2003) has shown that the five illocutionary classes are distributed with a specific probability of occurrence in informal spoken Italian (see the data below). Mapping on this statistics the percentages recorded in the Modal utterances corpus (Figure 3), that belong to both formal and informal, we can very roughly figure out the relative probability of each illocutionary class to bear modal indexes, which is much higher for the Representatives:

[5] In this research the annotation followed the definitions in the above references, which cannot be reported here. However, similar results would have been reached applying definitions of the illocutionary act as in Searle (1979).



Representatives *modalized*: ~ 18%
Directives *modalized*: ~7%
Expressives *modalized*: ~3,5%

Figure 3. Percentage of utterances of Illocutionary classes in a sampling of the LABLITA Corpus (Firenzuoli 2003)

The relation between the illocutionary classes and modal values is the main distributional evidence to the end of this paper. The following pies shows for each illocutionary class the relative percentage of modal types in Comment. One example for each modal type is reported under each pie[6].

Crucially there is no evidence that modal indexes select any specific illocutionary value and vice versa. On the contrary actual data shows that the utterances belonging to the main illocutionary classes can be accomplished with comment units bearing whatever modal value.

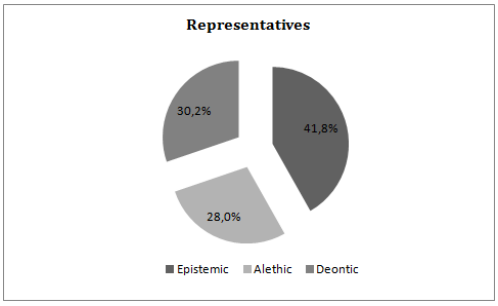


Figure 4. Types of Modal values in the Comment of Representatives utterances

Representative – Alethic

(11) *MAR: vedi /^{CON} adesso /^{TOP} i colori sono *sicuramente* questi //^{COM} [ifamcv09]
[Look /^{CON} at present /^{TOP} colors are *for sure* these ones ^{COM}]
%ill: Explanation

[6] To allow a better interpretation, the illocutionary type (%ill:) has been annotated after the translation. The reader can get prosody from the C-ORAL-ROM audio source.

Representative – Epistemic

(12) *ELA: quindi /[INP] *dovrebbe esserci* power point duemila //[COM] [iteldl02]
[*so* / [INP] *there should be* Power Point 2000 installed //[COM]]
%ill: Hypothesis

Representative – Deontic

(13) *ELE: *voglio* fare il trapianto //[COM] [imeddl02]
[I *want* the organ transplant // [COM]]
%ill: Expression of intention

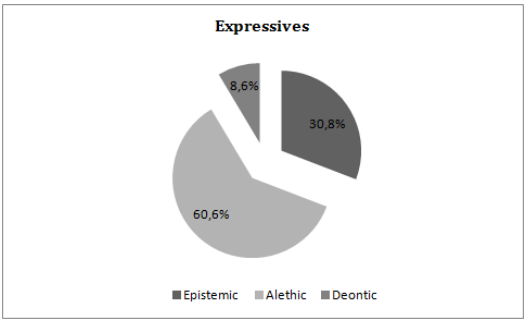


Figure 5. Types of Modal values in the Comment of Expressives utterances

Espressive – Alethic

(14) *ANG: cioè /[INP] sono *veramente* settemilalire //[COM] [ifamcv02]
[I mean /[INP] these are really 7,000 Liras //[COM]]
%ill: Expression of obviousness

Expressive – Epistemic

(15) *ROS: tu ce l' *avrai* te /[COM] i limiti di lingua //[APC] [ipubcv01]
[It is you that *might be* affected /[COM] by language limitations // [APC]]
%ill: taking offence

Espressive – Deontic

(16) *ANG: ma non te la *puoi* menare così /[COM] con la lunga scadenza //[APC] [ifamcv02]
[But you *cannot* go on this way / [COM] with the far dead line //[APC]]
%ill: reproaching

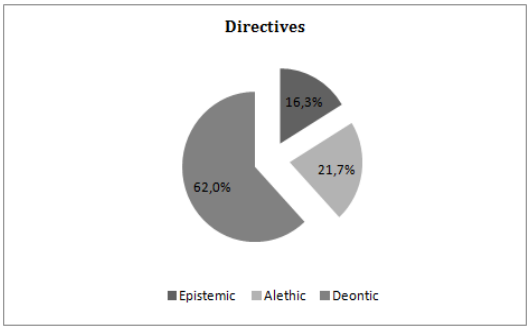


Figure 6. Types of Modal values in the Comment of Directives utterances

Directive – Deontic

(17) *DAV: *devi* mettere quella rossa //[COM] [ifamcv09]
[You *must* dress the red one //[COM]]
%ill: Instruction

Directive – Alethic

(18) *INA: *possiamo* vedere le immagini //[COM] [inednw01]
[We *can* sell images //[COM]]
%ill: Distal deixis

Directive – Epistemic

(19) *MAR: il primo incontro /[TOP] *credo* risalga ai suoi sedici anni ?[COM] [imedin01]
[the first meeting /[TOP] *I think* it was when she was sixteen? [COM]]
%ill: Request of confirmation

7. Conclusion

Summarizing, we have shown that the scope of modality in spontaneous speech is limited to the information unit boundaries and does not correspond to the scope of illocutionary force; that is the utterance.

There is a whole set of positive evidences supporting, in an harmonious way, that the scope of a modal value in speech can be considered the information unit: quantitative data of distribution, the fact that only specific types of information units can bear modal lexical indexes, the preference for specific modal values shown by

each type of information unit, and finally the impossibility of a compositional solution of modal values in different modalized information units. All these aspects can hardly been explained if a semantic entity such as the 'proposition of the utterance' is taken as the reference unit for modality in speech.

Moreover, given the complete reciprocal distribution of Modal types and Illocutionary classes the two notions are not a function the one of the other. Therefore, in no way modal indexes decides the illocutionary class. The two notions are definitively independent: Modality is a semantic aspect of the locutive program, in which the speaker's stance towards his locutory expression is manifested, while Illocutionary force belongs to pragmatics (the speaker manifests his attitude towards his interlocutor).

Beyond the limits of this specific issue, data shows however preferential correlations between the pragmatic aspect of illocutionary acts and the semantic of modal indexes. Some of these are obvious, i.e. Directive illocutions present a higher percentage of Deontic (62%). But others aspects are rather unexpected:

1. representative utterances present an high percentage of Epistemic values (41,8%);
2. the greater part of Expressive utterances in the corpus of Modal utterances are Alethic (60,6%). Both these finding are totally new and cannot be compared with any antecedent.

References

Austin, J.L. 1962. *How to Do Things With Words*. Oxford: Oxford University Press.

Bally, Ch. 1942. Syntaxe de la modalité explicite. *Cahiers Ferdinand de Saussure* 2: 3-13.

Bally, Ch. 1950 [1932]. *Linguistique générale et linguistique française*. Berna: Francke Verlag.

Bréal, M. 1924 [1897]. *Essai de Sémantique*. Paris: Hachette.

Bybee, J. & Fleischmann, S. (eds) 1995. *Modality and grammar in discourse*. Amsterdam/ Philadelphia: Benjamins.

Conte, A. 1977. Aspetti della semantica del linguaggio deontico. In G. Di Bernardo (ed.), *Logica deontica e semantica*. Bologna: Il Mulino.

Conte, M. E. 1995. Epistemico, deontico, anankastico. In A. Giacalone Ramat & G. Crocco Galèas (eds), *From Pragmatics to Syntax. Modality in Second Language Acquisition*. Tübingen: Narr, 3-9.

Cresti, E. 2000. *Corpus dell'italiano parlato.* Firenze: Accademia della Crusca.

Cresti, E. 2002. Illocuzione e modalità. In P. Beccaria, C. Marello (eds), *Scritti in onore di Bice Mortara-Garavelli*. Torino: Ed. Dell'Orso, 133-145.

Cresti, E. 2003. Illocution et modalité dans le comment et le topic. In Scarano, A. (ed.),

Macrosyntaxe et pragmatique: l'analyse linguistique del'oral. Roma: Bulzoni, 133-182.

Cresti, E. & Firenzuoli, V. 2001. Illocution et profils intonatifs de l'italien. *Revue Française de Linguistique Appliquée* 4, 2: 77-98.

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

Cresti, E. & Moneglia, M. 2006. C-ORAL-ROM. Prosodic Boundaries for Spontaneous Speech Analysis. In Y. Kawaguchi, et al. (eds), *Spoken Language Corpus and Linguistic Informatics*. Amsterdam: Benjamins, 89-113.

Diewald, G. 2001. A basic semantic template for lexical and grammaticalized uses of the german modals. In J. Van der Auwera & P. Dendale (eds), *Modal Verbs in Germanic and Romance Languages*. Amsterdam: Benjamins, 19-29.

Firenzuoli, V. 2003. Repertorio delle forme intonative di valore illocutivo dell'italiano. Analisi sperimentale di un corpus di parlato spontaneo (corpus LABLITA). PhD diss., Dipartimento di Linguistica, Università di Firenze.

Firenzuoli, V. & Signorini, S. 2003. L'unità informativa di topic: correlati intonativi. In G. Marotta & N. Nocch (eds), *Atti delle Giornate del Gruppo di Fonetica Sperimentale – XIII*. Pisa.

Frosali, F. 2008. Il Lessico degli ausili dialogici. In E. Cresti (ed.), *Prospettive nello studio del lessico italiano*, Atti del IX Congresso della Società Internazionale di Linguistica e Filologia Italiana (Firenze 14-17 giugno 2006). Firenze: FUP, 417-424.

Galvan, S. 1991. *Logiche Intensionali. Sistemi preposizionali di logica modale, deontica, epistemica*. Milano: Franco Angeli.

Graffi, G. 1994. *Sintassi*. Bologna: Il Mulino.

Green, M. 2007. Speech Acts. In E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/speech-acts/

Hare, R.M. 1961. *Il linguaggio della morale*. Roma: Ubaldini editore.

Hengeveld, K. 1988. Illocution, Mood and Modality in a Functional Grammar of Spanish. *Journal of Semantics* 6: 227-269.

Hoye, L. 1997. *Adverbs and Modality in English*. London: Longman.

Huddleston, R.D. 1984. *Introduction to the Grammar of English*. Cambridge: CUP.

Huddleston R. & G.K. Pullum, 2002. *The Cambridge Grammar of the English Language*. Cambridge: CUP.

Huges, G.E. & M.J. Cresswell, 1968. *An Introduction to Modal Logic*. London: Methuen.

James, F. 1986. *Semantics of the English Subjunctive.* Vancouver: University of British Columbia Press.

Kärkkäinen, E. 1987. Towards a Pragmatic Description of Modality. In I. Lindblad & M. Ljung (eds), *Proceedings from the Third Nordic Conference for English Studies*. Stockholm: Almqvist & Wicksell, 149-161.

Kiefer, F. 1994. Modality. In R.E. Asher (ed.), *The Encyclopedia of language and linguistics*. Oxford: Pergamon Press, 2515-2520.

Lyons, J. 1977. *Semantics*. 2 voll. Cambridge: CUP.

Mindt, D. 1995. *An empirical grammar of the English verb: modal verbs*. Berlin: Cornelson.

Mithun, M. 1995. On the Relativity of Irreality. In J. Bybee & S. Fleischman (eds), *Modality in Grammar and Discourse*. Amsterdam: Benjamins, 367-388.

Moneglia, M. 2011. Spoken Corpora and Pragmatics. In *Revista Brasileira de Linguística Aplicada.* http://www.scielo.br/scielo.php?pid=S1984-63982011000200009&script=sci_arttext

Palmer, F. R. 1990 [1986]. *Mood and Modality* 2nd ed. Cambridge: CUP.

Papafragou, A. 2001. *Modality: Issues in the Semantics-Pragmatics interface*. Amsterdam: Elsevier.

Perkins, M.R. 1983. *Modal expressions in English*. London: Frances Pinter.

Pottier, B. 2000. *Représentations mentales et catégorisations linguistiques*. Louvain/Paris: Peeters.

Schneider, S. 1999. *Il congiuntivo tra modalità e subordinazione*. Roma: Carocci.

Searle, J. 1979. *Expression and meaning: Studies in the theory of speech acts.* Cambridge: CUP.

Traugott, E.C. 1989. On the Rise of Epistemic Meanings in English: An example of Subjectification in Semantic Change. *Language* 65: 31-35.

Tucci, I. 2004. L'Inciso: caratteristiche morfosintattiche e intonative in un corpus di riferimento. In F. Albano Leoni et al. (eds), *Atti del Convegno Nazionale "Il Parlato Italiano"*. Napoli: M. D'Auria. (CD-ROM), 1-14.

Tucci, I. 2007. L'espressione lessicale della modalità nel parlato spontaneo. Dati dal corpus C-ORAL-ROM Italiano. PhD diss., Università di Firenze, Dipartimento di Linguistica.

Tucci, I. 2009. The scope of lexical modality and the informational structure in spoken Italian. In L. Mereu (ed.), *Information Structure and its Interfaces*. Berlin/NewYork: Mouton, 203-226.

Tucci, I. 2010. "Obiter dictum". La funzione informativa delle unità parentetiche. In F. Albano Leoni et al. (eds), *Atti del Convegno internazionale G.S.C.P. "La comunicazione parlata"*. Napoli: Università degli Studi di Napoli L'Orientale, vol. I, 635-654.

Venier, F. 1991. *La modalizzazione assertiva, avverbi modali e verbi parentetici.* Milano: Franco Angeli.

Vogeleer, S., Borillo A., Vuillaume M. & Vetters C. 1999. *La modalité sous Tous ses Aspects*. Amsterdam/Atlanta: Rodopi.

von Wright, G.H. 1951. *Essays on Modal Logic*. Amsterdam: North-Holland.

THE INVESTIGATION OF SPEECH EXPRESSIVITY

Sandra Madureira
Pontifical Catholic University of São Paulo (PUCSP)

1. Introduction

The objectives of this paper are threefold: considering theoretical issues concerning speech expressivity and sound symbolism; presenting the methodological procedures which have been developed in the investigation of speech expressivity at the Phonetics Laboratory (LIAAC) of the Pontifical Catholic University of São Paulo (PUCSP) and describing the results of applying these methodological procedures to the analysis of expressivity in a speech sample.

A research methodology comprising text interpretation (meaning production), prosodic perceptual analysis (intonation and rhythmical patterns and pause), prosodic acoustic analysis, perceptual analysis of voice qualities and perceptual analysis based on semantic descriptors are proposed.

For the analysis of voice qualities the adapted version (BP- VPAS) of the VPAS (Laver & Mackenzie-Beck 2007) has been used (See Appendix). According to the phonetic model of voice quality description (Laver 1980; Laver 2000) voice quality settings comprise both phonatory and articulatory adjustments from a neutral setting. The phonatory and articulatory settings modify the configurations of the speech tract and these changes yield specific acoustic outputs which influence listeners' judgments of paralinguistic features. The settings of raised larynx and spreading lips, for instance, in making the vocal tract smaller tend to rise fundamental frequency while lowered larynx and lip protrusion in enlarging the vocal tract tend to lower it. These changes affect listerners' judgments based on the frequency code (Ohala 1983; Ohala 1984; Chuenwattanapranithi 2008). The setting of phonatory or vocal tract tenseness are produced with greater muscular effort and tend to increase intensity and that affects listeners' judgments based on the effort code (Gussenhoven 2002; Gussenhoven 2004). The settings of creaky voice tends to occur at the end of utterances signaling finality. That has to do with the production code (Gussenhoven 2002; Gussenhoven 2004). In the production of creaky voice, vocal fold vibration rate diminishes and fundamental frequency is lower than in modal voice.

Mello H., Panunzi A., Raso T. (eds), *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation*

For the perceptual evaluation of the expressive uses of prosodic aspects, a group of listeners (judges) answers a semantic differential scale questionnaire having as descriptors emotional primitives (activation: calm/activated; valence: pleasant/ unpleasant and dominance (weak/strong), affective states (joy, sadness, anger, surprise) attitudes (aggressive, pleasant) or speech acts (advice, admonition, order and plea). For the acoustic analysis, based on PRAAT, manual and automatic measures (SG detector and SG Expressive Evaluator developed by Plínio Barbosa to analyze speech expressivity (Barbosa 2009) have been used and statistical measures calculated.

These methodological procedures have been taken into account in Madureira (2008) and Madureira & Camargo (2010). In Madureira (2008) speaking strategies used by two professional speakers, an actor and an actress) in reciting the poem *Soneto da Fidelidade* (Sonnet on Fidelity) were examined. Spectrographic and perceptual analysis of the recording of the sonnet were carried out. The speaking strategies used by the actors and actresses and their effects were contrasted to discuss relations between sound and meaning. The speakers' prosodic choices concerning voice quality settings, intonation patterns and distribution of pauses were found to differ and to affect the listener in dissimilar ways as shown by the results of the application of a semantic differential scale questionnaire to 30 judges. The actor's reading of the poem got the highest score for enthusiasm while that of the actress got the lowest score for that descriptor and the highest for sadness. Figure 1 displays the durational in ms and F0 contours in Hz of the sentence 'Mas que seja infinito enquanto dure' (*But might it be infinite while lasts*) of the poem *Soneto da Fidelidade* as produced by 6 speakers: three actors and three actresses, two of them analyzed in Madureira (2008).

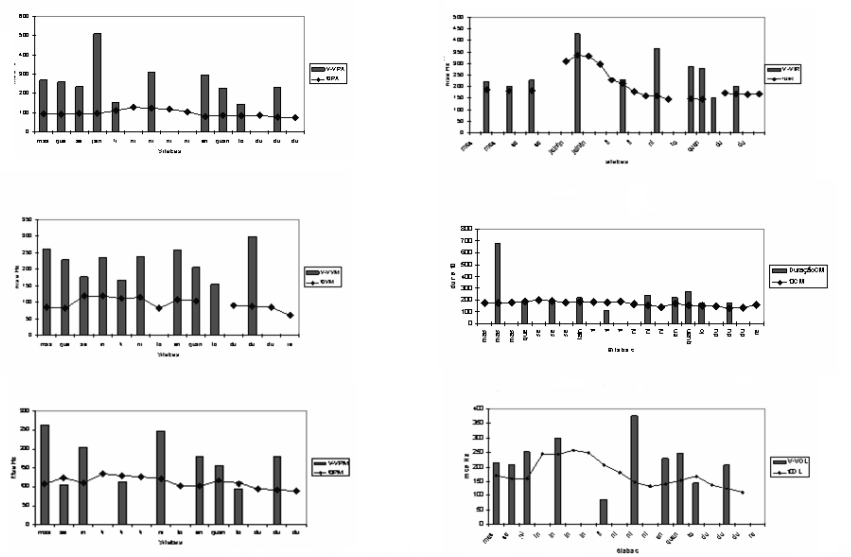


Figure 1. Duration values of V_V in ms (Colums) and F0 in Hz (Contours) of readings of the sentences 'Mas que seja infinito enquanto' dure by 6 speakers.

In Madureira & Camargo (2010) specific uses of sound symbolism concerning segmental and prosodic properties were examined in a reading of the poem A Valsa (The Waltz) by a professional actor. The typology developed by Hinton et al. (1994) was taken as reference. The results indicated the speaking strategies used by the actor have been found to make use of three types of sound symbolism (synesthesic, imitative and metalinguistic) to indicate both the dynamics of the dance as well as the dynamics of the conflicting affective states. Correlations among acoustic properties, perceived affective states and text meaning production demonstrate productive use of sound symbolism and corroborate the discussion on the direct links between sound and meaning.

The present paper takes into account the recordings of *A Valsa,* placing focus on the acoustic phonetic characteristics of the repeated stanzas and the methodological procedures used to analyze correlations between these characteristics and expressivity.

Reciting poems is a meaning-oriented production task. Metaphors are quite frequent in poems and some poetic narratives voices of various characters might be present. There is an aesthetic appeal to which the speaker has to respond. He is concerned with expressive ways of manifesting his interpretation of the text. His meaning production is influenced, among other factors, by his historical background, his knowledge of the themes being exploited, his affective conditions and the kind of

acting method he adopts. His prosodic choices affect the listeners' interpretations of his reading of the text.

2. Theoretical issues concerning speech expressivity

Beller (2009) defines speech expressivity as a level of information in communication. This level of information is referred to by Bolinger (1986) as deriving from the impressive potential of language. Adding to Beller's definition, the kind of information involved in speech expressivity is based on the interpretation of visual and vocal gestures and central to the discussion of the impressive effects of these gestures are matters of sound and meaning.

Barbosa (2009) presents a method combining two automatic acoustic analysis and multiple regression analysis for evaluating the degree of activation valence and involvement (emotional primitives) in speech expressivity.

Some of the key concepts related to speech expressivity are sound symbolism (Hinton et al. 1994; Ohala 1997) and sound metaphor (Fonagy 1983; Fonagy 2000). These concepts imply a functional direct link between sound and meaning. They have to do with form-function relations, which are based on three biological codes: the frequency code (Ohala 1983; Ohala 1984; Chuenwattanapranithi 2008), the production code and the effort code (Gussenhoven 2002; Gussenhoven 2004).

The frequency code is thought to have evolved from the animals' vocalizations in hostile situations (Morton 1977). The larger the animal the more aggressive it sounded. In speech, the correlations between larynx and vocal folds size and rate of vibrations of the vocal folds manifest power relations (strong/weak). Low pitch is associated with larger larynx and bigger vocal folds and can be used to signal strength and big things while high pitch is associated with smaller larynx and vocal folds and can be used to signal fragility and small things. Chuenwattanapranithi et al. (2006) report the findings of an experiment, which takes into account the dimension of size, and their conclusions corroborate the use of the size code to express emotions.

The effort code has to do with articulatory effort. The greater the articulatory effort the greater the tendency towards articulatory precision and greater prominence achieved by wider pitch range. The kinds of meanings which have been mentioned in the phonetic literature (Chen et al. 2002). to be associated with the effort cold are: emphasis, arousal, surprise.

The production code has to do with the generation of subglottal air pressure. At the beginning of utterances subglottal air pressure rises and at the end it declines. The kinds of meanings which might be associated with the production code are: continuity and finality.

The communicative power of the three biological codes as revealed by means of experiments provide evidence in favor of the close relation between sound and meaning.

In this paper a distinction between sound symbolism (the sound of meaning) and sound metaphor (the meaning of sound) is proposed. The expression 'sound symbolism' concerns the use of sound to produce meaning effects, that is, it refers to interpretations (meaning productions) based on some characteristics of the acoustic and physiological properties perceived by our senses and the expression sound metaphor refers to the choices of sound characteristics stemming from meaning productions and displaying some kind of analogy based on acoustic or physiological sensations.

A sentence as 'O relógio é dela' (*The clock is hers*) can be uttered to report someone's belonging without or with anger, sadness, joy, tenseness or any other kind of affective state being expressed. These feelings will be interpreted by listeners based on the acoustic characteristics of the speech production and that has to do with sound symbolism. Since an analogy between an affective state and the physiological conditions of voice production can also occur, a sound metaphor can also be derived. The muscular tension in the production of the utterance can yield an acoustic output and the meaning effect of psychological tension. Fónagy (2000: 345) argues that vocal gestures are metaphorical since they "imply transfer of a bodily gesture to the glottal or oral domain".

A sentence as 'O ritmo frenético do relógio' (*The frenetic rhythm of the clock*) said with a fast speech articulatory rate would be an instance of sound metaphor since an analogy is made between the rhythm of the clock and that of the speech production rate and it is that which motivates acoustic choices.

The distinction between sound symbolism and sound metaphor here proposed is based on the source and direction of the relation between sound and meaning: sound may produce meaning and meaning may produce sound (Albano 1988).

3. Methodological procedures in investigating speech expressivity in poetic corpora

The corpus of this work is a poem written in the nineteenth century by the Brazilian poet Casimiro de Abreu (1837-1860). It was recorded by a professional actor and the recording is available in a commercial CD entitled "Quatro Séculos de Poesia Brasileira" which was released by Luz da Cidade Productions in 2002.

Moraes (1989) presents an analysis of the rhythmical characteristics of this poem and concludes that the poetic structure metrics is based on the recurrent final tonic syllable occurring in regular intervals in the verse.

The poetic narrative takes into account the narrator's feelings towards his beloved one and his love rival while watching them dance. The poem has twenty three-syllable verses structured in eleven-line stanzas. One of the stanzas is repeated five times throughout the poem and although the syntactic and lexical items are the same, the affective states reported in the poem change throughout the text and affects their interpretation.

Choosing recordings of poems as speech research corpora enables the analysis of several kinds of interpretation: various speakers interpreting the same text; the same speaker interpreting various characters; the same speaker; reading repeated stanzas in different situational contexts depending on the affective states or social backgrounds being reported in the poem. The latter is the case of the poem "A Valsa".

The poetic narrative takes into account the narrator's feelings towards his beloved one and his rival while watching them dance. It comprises the dance compass, the dance dynamics (the speech rates changes from fast at the beginning of the poem to slow at the end) and the affective states dynamics (the narrator's feeling changes from love and admiration to jealousy, from exasperation to sadness).

For the purpose of this paper one of the stanzas of the poem, which is repeated five times, in the poem is considered. It comprises eleven verses:

Quem dera	(I hope)
Que sintas	(You feel)
As dores	(The pains)
De amores	(Out of love)
Que louco	(Crazy)
SentiI	(I've felt it)
Quem dera	(I hope)
Que sintas!...	(That you feel it!)
— Não negues	(Do not deny)
Não mintas...	(Do not lie …)
— Eu vi!	(I've seen it.)

The five repetitions, however, are preceded by stanzas whose informational structure are quite different. The first repetition occurs after a description of the physical beauty and attitudinal characteristics of the narrator's lover as she dances fast; the second after a stanza in which the narrator continues describing her lover's beauty, her attitudes and movements while dancing but manifests his jealousy, his exasperation and hate for his rival; the fourth follows a stanza in which the narrator's feelings of sadness are stated and the fifth follows the description of the

end of the dance, the end of the narrator's hopes and his lover's tiredness after the dance.

The ceasing of the dance/love and the tiredness/sadness feelings are described as *batida, caída, sem vida, no chão* (beaten, fallen lifelessly onto the ground). The actor uses voice dynamics and voice quality characteristics motivated by the semantic features of these lexical items, enhancing them. He speaks in a low speech rate, low pitch, producing a lowered larynx voice quality setting and introduces silent pauses.

In doing so he creates a sound metaphor based on the analogy between the lexical meaning features of the words and sound characteristics.

Acoustic measurement of F0 values taken at the speech sample 'chão' (*floor*) vary from 59 Hz to 78 Hz, contrasting with productions in which hate and jealousy is expressed. One of these occurs when, at some point in the poem the narrator asks: 'Mandavas a quem?' (*To whom do you address it (her smile)?)* In saying that utterance the narrator expresses his hate and jealousy towards his rival. That interpretation was corroborated by the results of an experiment in progress which consists of the application of a semantic differential scale questionnaire with the following descriptors (tenderness, cold anger, controlled anger, happiness and sadness) to 7 groups each of them containing repeated utterances and words taken from the poem, among them 'quem' (*who*). The judges, 30 university students from 20 to 30 years old judged the tokens extracted from the poem. The word 'quem' extracted from the utterance 'Mandavas a quem?' was judged to express to express anger, degrees varying in a 7 point scale. The results indicate choice of degrees 5 (20%), 6 (60%) and 7 (20%). The actor used a tense, hyper-articulated setting of voice, that is, the sound of his voice conveys some meaning effect which was not motivated by the lexical semantic features. The meaning effect has to do with his uses of voice dynamics and voice quality characteristics. It is a kind of sound symbolism.

Contrasting with the production of 'chão' (*floor*), the production of 'quem' (*whom*) in 'Mandavas a quem?' presents more variability and higher F0 values (from 179 Hz to 239 Hz). These findings are in accordance with findings in speech expressivity literature. Johnstone & Scherer (2000) report high F0 variability, high F0 values and wide F0 range among the acoustic correlates to anger. Figure displays the F0 contour of these two utterances. Figure displays the F0 contours of 'quem' and 'chão'.

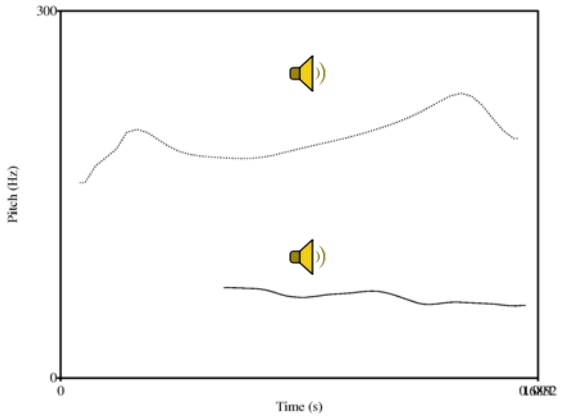


Figure 2. F0 contours of quem (upper contour) in the utterance Mandavas a quem and that of chão in the utterance caída no chão (lower contour).

Measurement of V-V units in ms of the five repetitions of the stanza chosen for the purpose of analysis in this paper were compared by means of ANOVA. No such differences were found among repetitions 1, 2 3 and 4. Repetition 5 was found to differ from the others p = 0.000.The fifth repetition was also found to differ from the others in relation to F0 (median, 99,5 quantil, skewness and its first derivate mean, standard deviation and skewness) and in relation to the long term average spectrum (LTAS). These differences in LTAS correlate with differences in voice quality identified by means of the VPAS. The fifth repetition is not produced with tenseness (neither laryngeal nor supralaryngeal) as the others are and its steep spectral slope as well as spectral characteristics in the frequency ranges 1-3 kHz and 4-5 kHz are compatible with the findings reported in Hammarberg & Gauffin (1995) and Nolan (1983) about the LTAS characteristics of the settings of hypofunction and whispery voice. The figures 3 and 4 present the LTAS curves and the trend lines of the five repetitions.

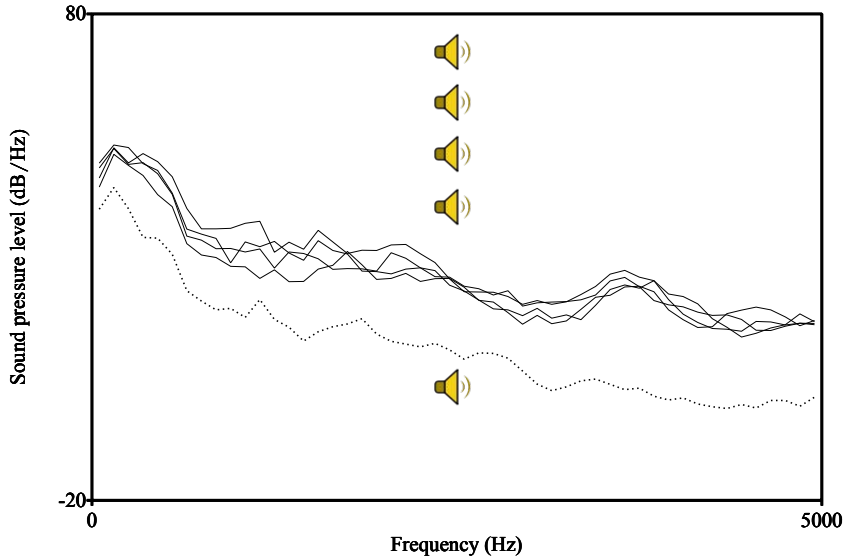


Figure 3. LTAS curves of the five repetitions of a stanza in the poem “A Valsa”. The dotted line refers to the fifth repetition.

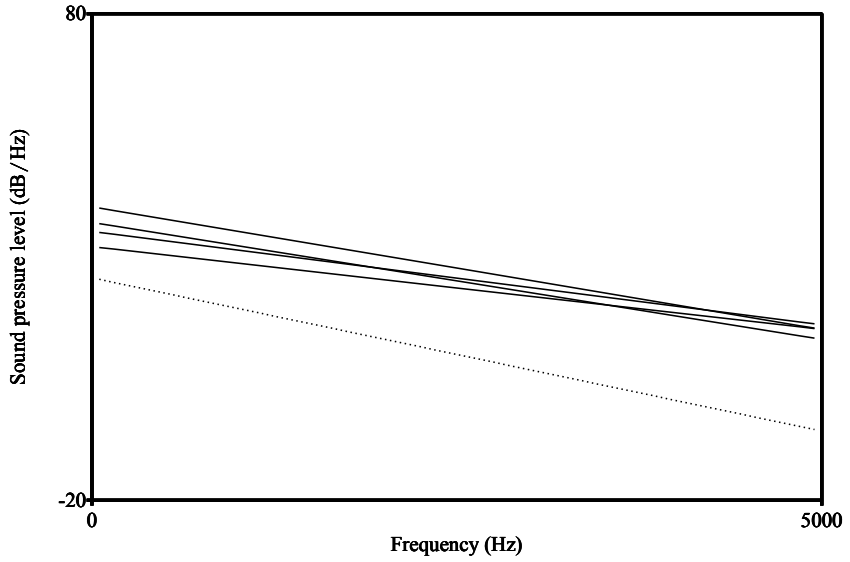


Figure 4. Trend lines related to LTAS curves of the five repetitions of a stanza in the poem “A Valsa”. The dotted line refers to the fifth repetition.

An acoustic and perceptual analysis of the three last verses 'Não negues' (*Do not deny*) 'Não mintas' (*Do not lie*); and 'Eu vi' (*I have seen it*) of the stanza repeated five times has also been carried out. The perceptual analysis comprised affective states (tenderness, exasperation, anger, happiness, sadness and fear) and speech acts (advice, plea threat, confirmation, request and admonition). These descriptors were included in a semantic differential rating scale questionnaire applied to a group of 30 judges.

The verses were produced with varied intonation patterns, pitch ranges and voice qualities throughout the text. There were also differences in F0 alignment and duration.

In the answers to the semantic differential scale questionnaire judges reported the second, the third and the fourth repetitions of 'Não mintas' (*Do not lie*), which were produced with hyperfunction (Tense Larynx) to be correlated to the expression of admonition, anger and exasperation and the first, which combined Tense and Raised Larynx, to request and the fifth, which was produced with expanded pharynx, to advice and plea The second, third and fourth repetitions of 'Não negues' were produced with Vocal Tract Tension and were correlated to threat and anger. The third repetition of 'Eu vi', which was produced with raised larynx was evaluated as indicating request and the first repetition which was produced with Tremor was evaluated as indicating fear and confirmation. The fifth repetition of 'Eu vi' was produced with whispery voice and evaluated as indicating tenderness.

Figures 5, 6 and 7 displays the waveform, the fundamental contour and the voice quality setting annotation of these three utterances.

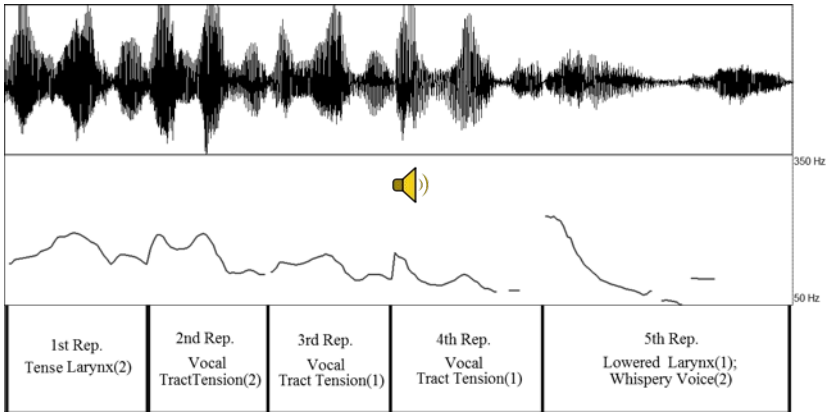


Figure 5. The waveform, the F0 contour, and a tier of annotation for the five repetitions of the utterance 'Não negues' (*Do not deny*). The number of the repetition and the type of voice quality setting are annotated.

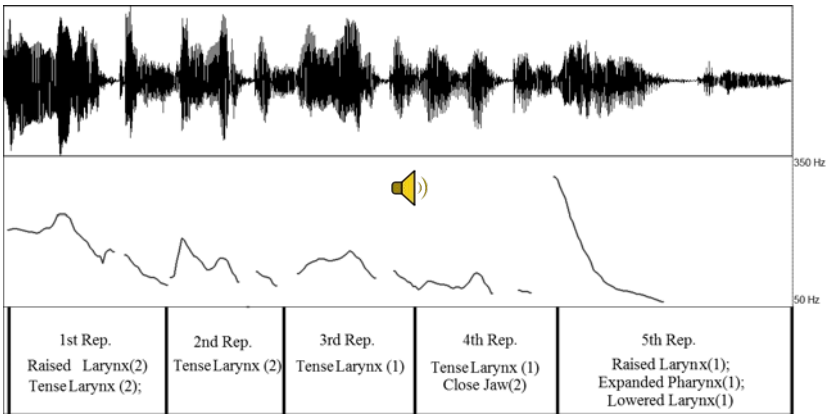


Figure 6. The waveform, the F0 contour, and a tier of annotation for the five repetitions of the utterance 'Não mintas' (*Do not lie*). The number of the repetition and the type of voice quality setting are annotated.

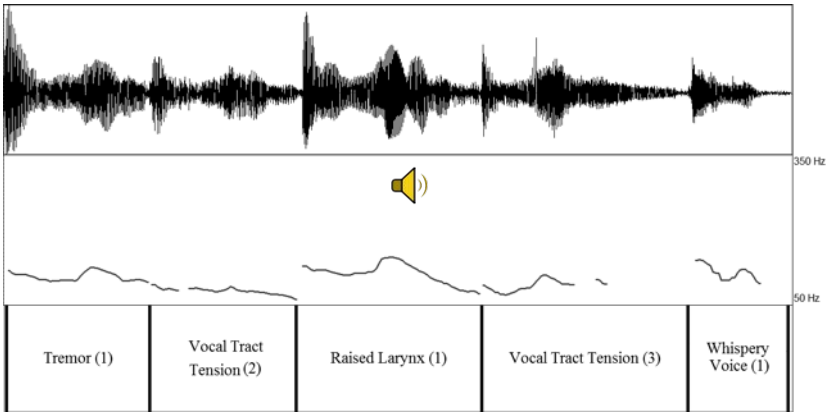


Figure 7. The waveform, the F0 contour, and a tier of annotation for the five repetitions of the utterance 'Eu vi' (*I have seen it*). The number of the repetition and the type of voice quality setting are annotated.

There is a cohesive prosodic relation among the four repetitions of the utterances 'Não negues' and 'Não mintas'. A declination line can be traced from the first repetition to the fourth. The pitch range gradually narrows and it can be interpreted as metaphorical representing the affective and dance dynamic changes that are reported throughout the poem. The fifth repetition of 'Não negues' 'Não mintas' exhibit a wider pitch range and a great fall in pitch which emphasizes the climax of the dynamics followed by the ceasing of the dance and that of love hopes. It signals

finality. There is also a cohesive relation between the first and the second and between the third and the fourth repetitions of the utterance 'Eu vi'.

4. Conclusions

Some correlations between voice quality settings and affective states can be thought of providing evidence in favor of the tenets of the frequency , production and effort codes. The voice quality setting of Raised Larynx, which tend to increase pitch, correlated with request but the utterances produced with Close Jaw and Vocal Tract Tension settings of voice quality were low in pitch and were found to signal threat. Those findings are compatible with the tenets of the frequency code. Larynx Tension settings imply in great muscular effort and correlated with admonition, exasperation and anger. On the contrary, Expanded Pharynx and Whispery Settings were found to correlate with tenderness, advice and plea and Tremor was correlated with fear.

The findings show that voice quality settings play an important role in speech expressivity and should be considered in combination of intonation and duration patterns. They are not only important to identify the kind of attitude or emotion but also the degree in which they are judged to manifest.

References

Albano, E.C. 1988. Fazendo sentido do Som. *Ilha do Desterro* 19: 11-26. Santa Catarina: Universidade Federal de Santa Catarina.

Barbosa, P.A. 2009. Detecting changes in speech expressiveness in participants of a radio program. In *Proceedings.of Interspeech 2009*. Brighton: United Kingdom, 2155-2158.

Beller, G. 2009. Transformation of Expressivity in Speech. In P. Lang (ed.), *The Role of Prosody in the Expression of Emotions in English and in French Peter Lang*. http://articles.ircam.fr/textes/Beller09c/index.pdf

Bolinger, D. 1986. *Intonation and its parts: melody in spoken English*. London: Edward Arnold.

Chen, A., Gussenhoven, C. & Rietveld, A. 2002. Language-specific Uses of the Effort Code. In B. Bel & I. Marlien (eds), *Proceedings of the 1st Conference on Speech Prosody*. Aix-en-Provence: Université de Provence, 215-218.

Chuenwattanapranithi, S., Xu, Y., Thipakorn, B. & Maneewonggvatana, S. 2008. Encoding emotions in speech with the size code: a perceptual investigation. *Phonetica* 65: 210-230.

Chuenwattanapranithi, S., Xu, Y., Thipakorn, B. & Maneewonggvatana, S. 2006. Expressing anger and joy with the size code. In *Proceedings of the 3rd tnternational Conference on Speech Prosody 2006*. Dresden, Germany.

Fonagy, I. 1983. *La vive voix*. Paris: Payot.

Fonagy, I. 2000. *Languages within languages: an evolutive approach*. Amsterdam-Philadelphia: John Benjamins Publishing Company.

Gussenhoven, C. 2002. Intonation and Interpretation: Phonetics and Phonology. In *Proceedings of the 1st International. Conference on Speech Prosody*. Aix-en-Provence, 47-57.

Gussenhoven, C. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Hammarberg, B. & Gauffin, J. 1995. Perceptual and acoustics characteristics of quality differences in pathological voices as related to physiological aspects. In O. Fujimura & M. Hirano (eds), *Vocal fold physiology*. San Diego: Singular Publishing, 283-303.

Hinton, L., Nichols, J. & Ohala, J.J. (eds) 1994. *Sound Symbolism*. Cambridge: Cambridge University Press.

Johnstone, T. & Scherer, K.R. 2000. Vocal Communication of Emotion. In M. Lewis & J. Haviland (eds), *Handbook of Emotions*. New York: Guilford Press, 220-23.

Laver, J. 1980. *The phonetic description of voice quality*. Cambridge: Cambridge University Press.

Laver, J. 2000. Phonetic evaluation of voice quality. In Kent R.D. & Ball M.J. (eds), *Voice quality measurement*. San Diego: Singular Thomson Learning, 37-48.

Laver, J. & Mackenzie-Beck J. 2007. Vocal profile Analysis Scheme – VPAS. Edinburgh: Queen Margareth University College - QMUC, Speech Science Research Centre.

Madureira, S. 2008. Reciting a sonnet: production strategies and perceptual effects. In *Proceedings of the Speech Prosody 2008 Conference*. São Paulo: Editora RG, 01, 697-700.

Madureira, S. & Camargo, Z.A. 2010. Exploring sound symbolism in the investigation of speech expressivity. In *Proceedings of ISCA, Tutorial and Research Workshop on Experimental Linguistics*. Athens, Greece, 105-108.

Moraes, J.A. 1989. Sobre o ritmo da poesia em língua portuguesa. Análise acústica das marcas prosódicas de fim de verso. In *Separata das Actas do XIX Congreso Internacional de Lingüística e Filoloxia Românica*. Coruña: Universidade de Santiago de Compostela, Fundación Pedro Barrié de La Maza Conde de Fenosa, 1017-1026.

Morton, E.W. 1977. On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist* 111: 855-869.

Nolan, F. 1983. *The phonetic bases of speaker recognition*. Cambridge: Cambridge University Press.

Ohala, J.J. 1997. Sound Symbolism. In *Proceedings of the 4th Seoul International Conference on Linguistics*. Seoul, Korea, 98-103.

Ohala, J.J. 1984. An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica* 41: 1-16.

Ohala, J.J. 1983. The origin of sound patterns in vocal tract constraints. In P.F. MacNeilage (ed.), *The Production of Speech*. New York: Springer-Verlag, 189-216.

Appendix

VPAS Laver & Mackenzie-Beck (2007)

Speaker: Date of recording: Judge: Recording ID:

	FIRST PASS		SECOND PASS						
	Neutral	Non-neutral	SETTING	Moderate			Extreme		
				1	2	3	4	5	6
A. VOCAL TRACT FEATURES									
1. Labial			Lip rounding/protrusion						
			Lip spreading						
			Labiodentalization						
			Minimised range						
			Extensive range						
2. Mandibular			Close jaw						
			Open jaw						
			Protruded jaw						
			Extensive range						
			Minimised range						
3. Lingual tip/blade			Advanced tip/blade						
			Retracted tip/blade						
4. Lingual body			Fronted tongue body						
			Backed tongue body						
			Raised tongue body						
			Lowered tongue body						
			Extensive range						
			Minimised range						
5. Pharyngeal			Pharyngeal constriction						
			Pharyngeal expansion						
6.Velopharyngeal			Audible nasal escape						
			Nasal						
			Denasal						
7. Larynx height			Raised Larynx						
			Lowered Larynx						

B. OVERALL MUSCULAR TENSION									
8. Vocal tract tension			Tense vocal tract						
			Lax vocal tract						
9. Laryngeal tension			Tense larynx						
			Lax larynx						

C. PHONATION FEATURES									
	SETTING	**Present**		**Scalar Degree**					
		Neutral	**Non-neutral**	**Moderate**			**Extreme**		
				1	**2**	**3**	**4**	**5**	**6**
10. Voicing type	Voice								
	Falsetto								
	Creak								
	Creaky								
11. Laryngeal frication	Whisper								
	Whispery								
12.Laryngeal irregularity	Harsh								
	Tremor								

	Neutral	**SETTING**	**Moderate**			**Extreme**		
			1	**2**	**3**	**4**	**5**	**6**
D. PROSODIC FEATURES								
13.Pitch	**Mean**	High						
		Low						
	Range	Minimised range						
		Extensive range						
	Variability	High						
		Low						
14. Loudness	**Mean**	High						
		Low						
	Range	Extensive range						
		Minimised range						
	Variability	High						
		Low						
E. TEMPORAL ORGANIZATION								
15. Continuity		Interrupted						

16. Rate		Fast						
		Slow						
F. OTHER FEATURES								
17. Respiratory support		Adequate						
		Inadequate						
18. Dyplophonia		Absent						
		Present						

BP - VPAS Camargo & Madureira (2008)

Nome: Data da gravação: Juiz: Identificação da gravação:

QUALIDADE VOCAL	**PRIMEIRA PASSADA**		**SEGUNDA PASSADA**						
	Neutro	**Não neutro**	**AJUSTE**	**Moderado**			**Extremo**		
				1	**2**	**3**	**4**	**5**	**6**
A. ELEMENTOS DO TRATO VOCAL									
1.Lábios			Arredondados/protraídos						
			Estirados						
			Labiodentalização						
			Extensão diminuída						
			Extensão aumentada						
2. Mandíbula			Fechada						
			Aberta						
			Protraída						
			Extensão diminuída						
			Extensão aumentada						
3. Língua ponta/lâmina			Avançada						
			Recuada						

4. Corpo de língua			Avançado						
			Recuado						
			Elevado						
			Abaixado						
			Extensão diminuída						
			Extensão aumentada						
5. Faringe			Constrição						
			Expansão						
6.Velofaringe			Escape nasal audível						
			Nasal						
			Denasal						
7. Altura de laringe			Elevada						
			Abaixada						
B. TENSÃO MUSCULAR GERAL									
8. Tensão do trato vocal			Hiperfunção						
			Hipofunção						
9. Tensão laríngea			Hiperfunção						
			Hipofunção						

C. ELEMENTOS FONATÓRIOS									
	AJUSTE	**Presente**		**Graus de escala**					
		Neutro	**Não Neutro**	**Moderado**			**Extremo**		
				1	**2**	**3**	**4**	**5**	**6**
10. Modo de fonação	Modal								
	Falsete								
	Crepitância/ vocal fry								
	Voz crepitante								
11. Fricção laríngea	Escape de ar								
	Voz soprosa								
12.Irregularidade laríngea	Voz áspera								

Ocorrências em curto termo ( ) quebras ( ) instabilidades ( ) diplofonia ( ) tremor

DINÂMICA VOCAL		**Neutro**	**AJUSTE**	**Moderado**			**Extremo**		
				1	**2**	**3**	**4**	**5**	**6**
D. ELEMENTOS PROSÓDICOS									
13. Pitch (F0)	Habitual		Elevado						
			Abaixado						
	Extensão		Diminuída						
			Aumentada						
	Variabilidade		Diminuída						
			Aumentada						
14. Loudness (intensidade)	Habitual		Aumentado						
			Diminuído						
	Extensão		Diminuída						
			Aumentada						
	Variabilidade		Diminuída						
			Aumentada						
15. Tempo									
Continuidade			Interrompida						
Taxa de elocução			Rápida						
			Lenta						
16. Outros Elementos									
Suporte respiratório			Adequado						
			Inadequado						
			Presente						

Para ajustes de ocorrência intermitente assinalar (i)

SPEECH RHYTHM AS A PATH BETWEEN STRUCTURING AND REGULARITY

AN OPTIMAL SOLUTION DURING THE ACT OF COMMUNICATING

Plínio A. Barbosa

Linguistics Department/IEL/State University of Campinas, Brazil

1. Introduction

This study is carried out in such a way as to explore formal devices for answering to this question: what makes utterances sound prosodically distinct in different speakers, in different speaking styles and in different language varieties? The first and second differences are usually investigated by the area of Stylistics, while the latter by the area of Typology. Rhythm is the domain of prosody chosen to tackle these stylistic and typological problems because we think speech rhythm is mainly what is modified in these distinct conditions of speech production. The methodology for stydying speech rhythm is that given by coupled-oscillator theories, because these theories are able to deal with the hierarchical structure of speech timing. This paper also aims at showing the advantages of coupled-oscillator theories to reveal speech rhythm patterns, while tooking the position the afore-mentionned differences can be mainly attributed to rhythm.

The experimental psychologist Paul Fraisse considered all rhythms as a result of two interacting components. This position could be summarised as "la structure se trouve toujours coulée dans une périodicité et la périodicité est toujours organisation de structure" (Fraisse 1968: 28). As Sauvanet (2000: 160-162) reminded us, Paul Valéry took the same position. In his obsession about how to define rhythm, he insisted on its inequivalence to periodicity or to regularity as early as in 1915:

> il ne faut pas mêler et encore moins confondre, *période* et *rythme*. Il n'est pas exact de dire : rythme des flots, rythme du coeur – etc. Ce sont des faits périodiques, *si l'on veut*. (Valéry 1973 [1915]: 1282)

Mello H., Panunzi A., Raso T. (eds), *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation*

The mathematician Whitehead (1919: 198), who said that "the essence of rhythm is the fusion of sameness and novelty", have already pointed out this ambivalent nature of rhythm, which includes that of linguistic rhythm. If rhythm were equationated with regularity alone it would be irrelevant for perception, because our attentional mechanisms seek for novelty (Cowan 1997: 149-151).

In these lines, Barbosa (2006) showed the advantage of recovering the position taken by Fraisse with the proposal of a dynamical model of speech rhythm (henceforth DRM). The computational implementation of this model is couched in dynamical systems theory (Kelso 1995), and presupposes that the rhythmic system underlying speech communication can be modelled by the coupling between two components, a perception-oriented component related to pattern structuring, and a production-oriented component operating under regularity constraints.

The first component of the DRM takes into account the coupling (reciprocal influence) between local syntactic information and a phrase-stress oscillator, while the second presupposes the coupling of two subcomponents, a syllabic oscillator and a phrase-stress oscillator, parameterised by a coupling strength.

The first level of coupling was implemented by a likelihood function defined within a window containing three putative prosodic boundaries after the corresponding three phonological words (Barbosa 2007). This function combines the probability of assigning a prosodic boundary given the strength of the local syntactic cohesion at each one of the three phonological words' boundary (syntactic constraint), with the probability of assigning a prosodic boundary given the distance in number of syllables since the last assigned boundary (regularity constraint). The two probabilities are linearly and complementarily combined by a parameter which rules the degree of influence of syntax (and complementarily, of constraints of regularity) in defining each prosodic boundary. This implementation generates both position and strength of prosodic boundary within the window.

The regularity of both oscillators in the second component generates complex patterns of syllable-sized durations as a consequence of the phrase-stress oscillator's influence onto the syllabic oscillator, under the guidance of the specifications of position and strength of prosodic boundary given by the first component.

Speech rate, specified underlyingly by the inverse of the syllabic oscillator resting period, is a basic property of the model. This means that syllable rate (which is the inverse of syllable duration mean) is only strictly periodic at the underlying level, and not at the surface, where syllable duration varies according to a great number of contextual variables.

Which is important to stress with relation to the DRM for the purpose of this paper is that the natural variation of syllable duration it delivers is a consequence of the interaction of regularity components at different paces which are coupled with each other. In the framework of the DRM, distinct speaking styles and different linguistic rhythms are a consequence of changes in the way this interaction takes

place. The investigation of how the model would work in these particular circumstances is the theme of this paper. There is at least one advantage of modelling over description of rhythm: the former allows predicting the behaviour of duration patterning for situations which were not previously described, and, by doing so, shed some light on the possible components of speech rhythm as well as on the way these components interact with each other.

1.1 The core of the DRM: the coupled-oscillator component

O'Dell and Nieminen (1999) proposed a simple way to infer the coupling strength parameter value which specifies the magnitude of the mutual influence between a syllabic oscillator and a phrase-stress oscillator. The Averaged Phase Difference technique is applied to infer the coupling strength value, provided that two conditions be satisfied: (a) that the coupling forces in both directions are symmetrical and differing only in sign and in the coupling strength of the phrase-stress oscillator onto the syllable oscillator, and (b) that the consequences of the coupling for both oscillators derive solely from these bidirectional forces, and from the number of cycles of the faster oscillator within the cycle of the slower oscillator. The authors showed that the coupling strength r between the two oscillators is equal to the ratio between the intersect, $r/(r.\ \omega_{ps} + \omega_{\sigma})$, and the slope, $1/(r.\ \omega_{ps} + \omega_{\sigma})$, of the linear regression computed between two variables: I and n in equation (1). Note that the ratio intersect/slope is equal to r. In equation (1), I is the duration of the stress group; ω_{ps} is the frequency of the stress group oscillator; ω_{σ}, the frequency of the syllabic oscillator; n is the number of syllables within the stress group; and $H(\Phi_n)$ the coupling function.

$$(1) \qquad I = \frac{1}{\omega_{ps} + H(\Phi n)} = \frac{r}{r.\omega_{ps} + \omega_{\sigma}} + n.\frac{1}{r.\omega_{ps} + \omega_{\sigma}}$$

This proposal represents a paradigm change in speech rhythm research, because it allowed restating early analysis on isochrony in relative terms: the higher the coupling strength, the more stress-timed a language is, and vice-versa. There is no need to refer to any kind of absolute isochrony. Indeed, provided that both regression coefficients are significant, if r = 1, this stands for an even influence between both oscillators. On the other hand, if $0 < r < 1$, the syllabic oscillator dominates the phrase stress oscillator (syllable timing), and if $r > 1$, the phrase stress oscillator dominates the syllabic oscillator (stress timing). Distinct languages or varieties, as well as speakers and speaking styles would differ in degree of coupling, but not in nature of the underlying phenomenon.

In order to find r, what is necessary is the computation of a linear regression having as dependent variable the duration of the stress group, and as explanatory

variable the number of syllables within it. To separate the contribution of this latter variable from effects of prosodic strength, O'Dell and Nieminen's proposal is modified here to include an estimation of prosodic strength as an additional explanatory variable. This decision has to do with differences in the treatment of coupled oscillators in cases of more than two levels of the prosodic hierarchy.

Recently, O'Dell et al. (2008) treated this issue by introducing additional oscillators with distinct periods, and then dealing with all levels of interaction between them, which adds to complexity. The DRM treats distinct levels in the prosody hierarchy by coding these levels in terms of magnitude of the pulses of the phrase stress oscillator, and not in terms of period. This allows a simplification in modelling, but requires the introduction of an additional explanatory variable that factor out these other levels of prosody information. This additional variable stands for prosodic boundary strength (z^n_{sm}) and is presented in section 2.3.

2. Methodology

To approach the issue of characterising distinct language varieties rhythmwise, Brazilian (henceforth BP) and European Portuguese (henceforth EP) were chosen. The reason for this choice is related to the allegedly prosodic differences (Frota et al. 2002) between the two varieties. That this difference is not an illusion of other linguistic and paralinguistic aspects of these varieties and can be partly attributed to rhythm alone will be shown in the next section.

For evaluating possible rhythmic differences in terms of speaking style, reading vs storytelling styles were chosen. This choice is motivated by the fact that storytelling presents elements that can be found in spontaneous conversation, such as hesitations due to macro- and microplanning of the discourse. Though read speech can have hesitations, these are much lesser frequent than in the case of storytelling. This feature is important to approach a description of speech rhythm in natural conditions and investigate the possible differences between less and more controlled situations of utterance production.

Different speakers in each speaking style and language variety were chosen in order to evaluate in what respects people could differ in terms of rhythm, at least as modelled by the DRM.

Following the interplay of regularity and structuring, the variables chosen for analysis were stress group duration (I), number of syllables (n) within the stress group, and a measure of prosodic strength (z^n_{sm}) at the right edge of the stress group. The choice of these variables aims at investigating the respective roles of the regular sucession of syllables, and of prosodic boundary strength for explaining stress group duration. Stress group is a unit which has one prominent syllable preceded, in the case of the varieties studied here, by a variable number of non prominent syllables.

This prominent syllable bears the so-called phrase stress. The appropriate statistical technique to enquiry about these relations is multiple linear regression. This analysis was made by using the R statistical package (R project).

2.1 Corpora

The corpora consisted of parallel productions of six subjects in both EP and BP. Two native female and one native male speakers for each variety read a 1,400-word text on the origin of the pastries *pastéis de Belém* (reading style, RE). After the reading, the six subjects told what the text was about (story telling style, ST). Each native speaker read the text written in his/her own written variety. All speakers aged 30 to 45 years, and were full or student researchers on speech science and technology. As the stories told by some speakers was much shorter than the reading material, excerpts containing about 350 words were chosen for analysis in the twelve productions (six speakers and two speaking styles), with the exception of the Brazilian male speaker, who told the story in 141 words..

2.2 Measured variables and techniques of analysis

Following a traditional approach in speech research (cf. Classe 1939; Lehiste 1970; Dogil & Braun 1988, *inter alia*), syllables were phonetically segmented by tracking two consecutive vowel onsets (VO). These points were marked semi-automatically in Praat (Boersma & Weenink 2008) into two stages: automatic VO detection by the *Beatextractor* Praat script (Barbosa 2006) followed by manual correction, where applicable. This script detects points in the speech signal where changes in previously filtered energy envelope are relatively fast and positive (from low to high energy). According to Scott's (1993) work, the speech signal energy was filtered in the region of the first and second formants to simulate the way our auditory system works for detecting syllables.

Each interval delimited that way defines a VV unit with a specified duration computed automatically from the segmented speech signal. More than 3,450 VV units were segmented and manually tagged with a broad phonetic transcription. Stress groups were delimited by automatically detecting phrase stress boundaries within and across connected utterances. Because syllable-sized duration is a main parameter specifying both lexical and phrase stress in Portuguese (for BP, see Massini 1991; Barbosa 1996), normalised VV durations were chosen as a measure of prosodic strength to detect phrase stress position.

The sequence of phrase stress positions was then automatically tracked by serially applying two techniques for normalising the VV durations. The first one was a z − score transform applied to each VV unit *i*:

$$z^i = \frac{\mathrm{dur} - \sum_j \mu_j}{\sqrt{\sum_j \mathrm{var}_j}} \quad (2)$$

In (2), dur is the VV duration in milliseconds, whereas the pair (μ_j, var_j) are the reference mean and variance in milliseconds of the segments within the corresponding VV unit. These reference values are found in Barbosa (2006: 489) for BP. For EP, a reference table was created from the analysis of a corpus of read speech in a project held by the INESC-Lisbon. The transformation in (2) was followed by a 5-point moving average filtering in (3), where z^i_{sm} is the smoothed value of z for the i^{th} VV unit.

$$z^i_{sm} = \frac{5.z^i + 3.z^{i-1} + 3.z^{i+1} + 1.z^{i-2} + 1.z^{i+2}}{13} \quad (3)$$

In both BP and EP phrase stress is placed at the right edge of the duration-related stress group. The normalisation technique above, followed by the automatic tracking of duration-related phrase stress boundaries from the detection of smoothed z^i maxima were implemented by a Praat script (SGdetector, available from the author). The computation of both the stress group duration and the number of VV units in the stress group is also done by the SGdetector script. The number of phonological syllables was computed manually for each stress group. They will be referred to here simply as syllables.

Since the procedure of stress group segmentation is entirely based on duration maxima, the right boundary not necessarily coincides with a lexically stressed unit. Sometimes a post-stressed lengthened VV unit signals the end of the stress group. Silent pauses were included in the VV units that precede them. In doing so, high values of z^i_{sm} were obtained from VV units containing silent pauses, signalling a strong prosodic boundary. Offglides were included in the VV unit containing the vowel leftwards. Onglides formed a vocalic unit with the vowel rightwards.

The values z^n_{sm} stand for the measure of prosodic strength at the end of each stress group (of size n). They allow to factoring out levels higher than two in the prosody hierarchical, leaving the possibility of examining the relation between only two levels of oscillation, the syllabic and the phrase stress oscillations of the 1999 O'Dell and Nieminen proposal. The ratio between the intersect and the coefficient associated to the number of syllable-sized unit is the estimation of the coupling strength r. Differences in coupling strength would reflect differences in speaking style, speaker, and language variety, which can be studied statistically.

An example of the application of the two-stage technique of VV duration normalisation will be illustrated in the following. Figure 1 shows the values of raw (non-normalised) durations for the VV units of the sentence "Manuel tinha entrado apra o mosterio há quase um ano, mas ainda não se adaptara àquela maneira de viver." read by the Brazilian female speaker LL. The y-axis shows each VV duration

value in milliseconds, while de x-axis shows part of the VV units of the corresponding utterance. Transcriptions are made using the I.P.A. with capital letters standing for archiphonemes. Each VV units starts and ends at a vowel onset. The first one starts at the onset of the vowel /a/ of the word 'Manuel' and end at the onset of the vowel /u/ in the next syllable, which gives the unit /an/. For the reasons explained above, the third VV unit, /iNtɾ/ from the sequence 't*i(nha) entr*ado' is formed that way because of the deletion of the palatal nasal and of the final /a/ of 'tinha', as well as the sândhi between /i/ of 'tinha' (pronounced [ĩ]) and /en/ of 'entrado' (pronounced likewise).

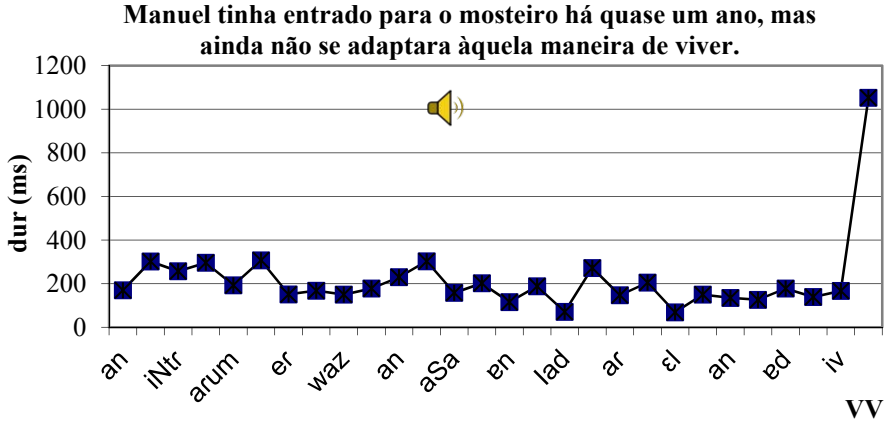


Figure 1. Values of raw (non-normalised) VV durations for the sentence "Manuel tinha entrado para o mosteiro há quase um ano, mas ainda não se adaptara àquela maneira de viver.", uttered by the BP female speaker LL.

Observe at least eleven local peaks of VV duration in Figure 1. Not all of these peaks are perceived as salient by a listener. The normalisation technique aims at making salient the VV units which are likely to be perceived as prominents by a listener. The application of the two steps presented above gives the patterns shown in Figure 2.

Still a considerable amount of local peaks persist after the application of equation 2 (diamonds in the figure), although a secondary peak emerges from he durational pattern very clearly now, that at the end of the word 'ano' at the strongest syntactic (and prosodic break), between the two coordinated clauses.

The application of equation 3 allows to confirm the two strongest boundaries after 'ano', and at the end of the utterance. The exam of the values of smoothed z-scores reveal three additional weaker boundaries after 'tinha', 'mosteiro' and 'adaptara' (with respective values of -0.06, -0.12, -0.84), which correspond closely to the general perception of where boundaries and prominences are in this example.

In our approach, as explained above, these four boundaries define four phrase stresses at the preceding words, with different degrees of strength. Each one of the four values of strength in the example shown here is given by the value of the four smoothed z-score local peaks.

With the triads duration of stress group (I), number of syllables (or VV units) within it (n), and smoothed z-score of the last VV unit in the corresponding stress group (z^n_{sm}) at hand, the following multiple linear regression was computed for number of syllables and number of VV units in the stress group:

(4) $$I = a + b.n + c.z^n_{sm}$$

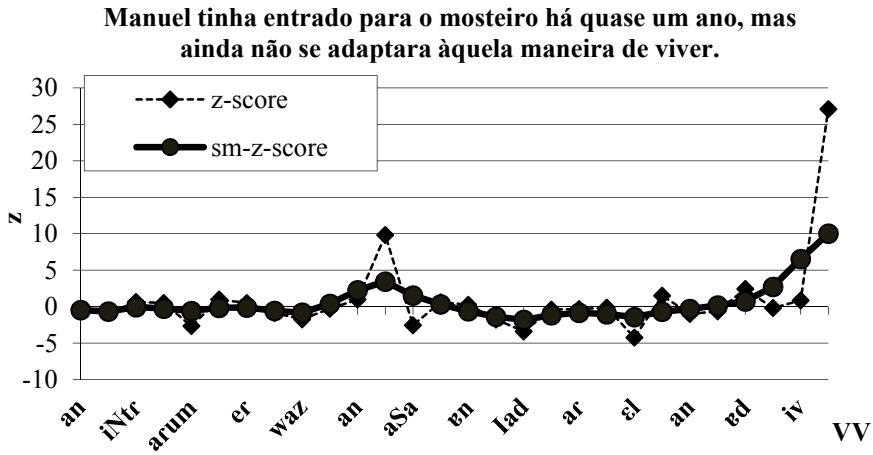


Figure 2. Values of normalised VV durations for the sentence "Manuel tinha entrado para o mosteiro há quase um ano, mas ainda não se adaptara àquela maneira de viver.", uttered by the BP female speaker LL. Diamons connect with dashed lines show z-score, wheras filled circles show smmothed z-scores.

3. Results

As the number of VV units in the stress group did not turn out to produce significant values for some intersect coefficients, only the linear regressions taking the number n of syllables as an explanatory variable are shown here. Table 1 shows the linear regression equations, according to language variety (BP/EP), speaker and sex (LLF stands for speaker LL, female, for instance), as well as speaking style (RE or ST).

The coupling strength r in Table 1 is the ratio between the intersect and the slope coefficient for the number n of syllables in the respective equation. All correlation coefficients and (consequently) the inclination coefficients for both

explanatory variables are highly significant ($p < 10^{-4}$). The significance of the intersect coefficient is indicated between parentheses. In case of non significant values for this coefficient, r is considered undefined (u), except in cases of marginal significance. Speech rate (sr) is given in syllables/s. All correlation coefficients are between 0.79 and 0.97. These figures mean that 62 % to 94 % of the variance of stress group duration is explained from the number of syllables combined with the estimated prosodic boundary strength given by smoothed z peaks. The analysis of the multiple regression reveals that explanatory variables, n and z^n_{sm}, contribute independently to predict I (cross-variable R^2 is inferior to 0.003 for all cases).

Table 1: Multiple regression equations for two language varieties (BP and EP), six speakers (LL, AG, FA, SV, AJ, and IT), and two speaking styles (ST and RE). The letter after the subject label stands for female (F) and male (M). The significance of the intersect parameter a is given in the 3rd column. Coupling strength r is computed for significant values of parameter a. I stands for stress group duration, n, for the number of phonological syllables, and z for phrase stress magnitude. Speech rate (sr) is expressed in syllables per second.

var-sp-sty	**equation**	**a signif**	**r**	**sr**
BP-LLF-RE	I = 215 + 126.n + 63.z	p<0.02	1.7	5.1
BP-LLF-ST	I = −10 + 182.n + 45.z	ns	u.	4.2
BP-AGF-RE	I = 71 + 153.n + 62.z	ns	u.	4.9
BP-AGF-ST	I = 373 + 138.n + 41.z	p < 0.005	2.7	4.1
BP-FAM-RE	I = 197 + 125.n + 63.z	p < 0.05	1.6	5.4
BP-FAM-ST	I = 237 + 143.n + 45.z	p < 0.07	1.7	4.3
EP-SVF-RE	I = 128 + 131.n + 156.z	p < 0.06	2.0	5.1
EP-SVF-ST	I = 441 + 124.n + 78.z	$p < 10^{-2}$	1.0	4.7
EP-AJM-RE	I = 103 + 126.n + 145.z	p < 0.1	0.8	6.2
EP-AJM-ST	I = 319 + 101.n + 131.z	$p < 10^{-2}$	3.2	5.7
EP-ITF-RE	I = 79 + 135.n + 161.z	ns	u.	5.3
EP-ITF-ST	I = 346 + 124.n + 109.z	$p < 10^{-2}$	2.8	4.8

The results in Table 1 show that speech rate is distinct from coupling strength: faster rates do not correspond necessarily to higher values for coupling strength, as can be seen for the male Portuguese speaker AJ in Reading style: he is the fastest speaker but his reading does not have the highest value of r (in fact, the value of r is close to the one of the female Portuguese speaker SV in storytelling style, who utters at a much lower speech rate). Portuguese speakers tend to have higher values of r in the storytelling style than Brazilian speakers. Higher values of coupling strength mean that storytelling is more stress-timed in EP than in BP. In the reading style, both varieties are very close in terms of coupling strength. Compare the similarity of the use of prominence and boundary in the reading of EP speaker SV with that of BP

speaker LL, both females [LLRE, SVRE]. Compare also the distinction in terms of the use of prominence and boundary in storytelling vs reading styles in EP speaker AJ [AJRE, AJST].

As signalled above, these equations were obtained by using number of syllables in the stress group as one of the explanatory variables. Although the computation of this number can be made automatic with a device such as an aligner (cf. lingWAVES; Goldman 2007), the use of VV units over syllables has the advantage of allowing the task of obtaining coupling strength values fully automatic. For doing so, it is necessary to avoid the stage of manual tagging of VV units with a phonetic label before duration normalisation. This was recently proposed by Barbosa (2010) and is currently under full testing.

The coupling strength values can be compared in terms of statistical significance too. What is needed is to compare the significance of the differences of the equations' parameters by using the ANCOVA technique. For illustrating this technique with number of VV units in the stress group as explanatory variable, the regression lines for Brazilian speakers AG and FA in the reading style were compared. Data and regression lines for the relation between number of VV units in the stress group (n) and duration of stress groups (DurSG) can be seen in Figure 3. Data from the entire reading of the corpus by both speakers was used in this illustration.

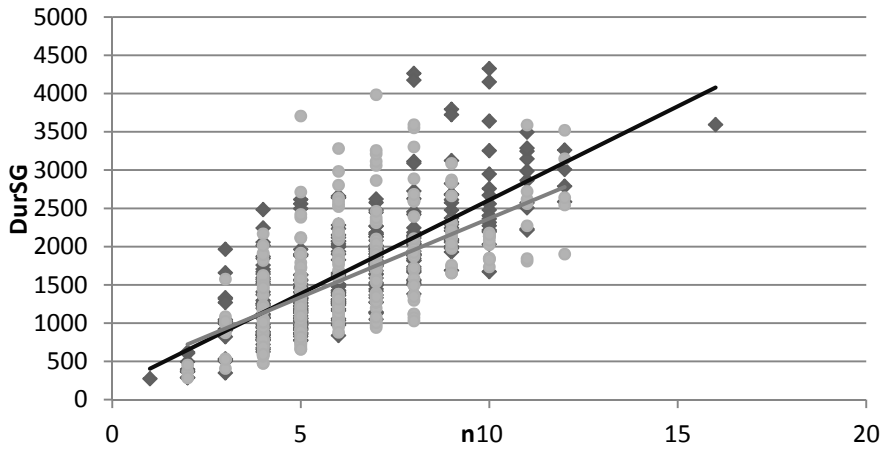


Figure 3. Data and regression lines for duration of stress group (DurSG) against number of VV units (n) in the reading style for Brazilian male FA (light gray circles) and female AG (dark gray diamonds). Observe the less steeper slope for the male speaker.

From the application of the ANCOVA the following equations were obtained:

(5) $\mathrm{I\,(AG)} = 0 + 220.\mathrm{n} + 43.z_{sm}^{n}$, and $\mathrm{I\,(FA)} = 165 + 180.\mathrm{n} + 56.z_{sm}^{n}$

The intersects are significantly marginally distinct from each other ($p < 0.09$), although the intersect coefficient of data from speaker AG is not distinct from zero. The coefficients of the n parameters are significantly distinct from each other ($p < 0.002$) and from zero ($p < 10^{-15}$), as well as the coefficients of the z^{n}_{sm} parameters from each other ($p < 0.0003$) and from zero ($p < 10^{-15}$). The combined data for the two explanatory variables explain 74 % of the variance of stress group duration. These figures give a coupling strength of zero (non significant) for AG data, distinct from the value of 0.92 ($r = 165/180$) for FA data. Speaker FA is then more stress-timed than speaker AG when reading [FARE, AGRE]. Compare the excerpts of their readings by paying attention to the more variable, more performed way speaker FA marks prominence and boundaries in comparison with speaker AG.

4. Discussion

As it was shown throughout the previous section, the coupling strength parameter seems to reflect differences in the subjects' rhythmic performance not only across language varieties (see Table 1 and *infra*), but also across speaking styles (see Table 1 and *infra*) and across individuals (see analysis of data of Figure 3). This picture gives, then, a partial answer to the question formulated in the beginning of this paper, what makes speech sound prosodically distinct in different speakers, in different speaking styles and in different language varieties is possibly the way the individuals manage to couple the production of syllables with the activity of structuring prominences and boundaries in specific situations of discourse, inside a particular linguistic community.

The DRM is a framework for studying the variation of syllable-sized duration pattern. As seen in the Introduction, it allows a way of explaining duration pattern complexity from two simple universal oscillators in interaction. The model actually produces surface duration as demonstrated in previous work (Barbosa 2007), and is also able to deal with secondary stress (Arantes 2010).

The analyses shown here can be rendered fully automatic, which enables the techniques presented to be used in the automatic identification of rhythm differences. A first step in this direction was presented recently (Barbosa 2010). It important to emphasise that the techniques presented are able to signal statistically significant differences between durational patterning between a set of utterances for a same or different speakers in particular situations and from possible distinct languages, separating prosodic from segmental structure. It is not possible to affect a particular excerpt of speech to a particular rhythm type, but only differences. We see this impossibility as an advantage, and not a drawback.

Both universal and language-specific aspects of speech rhythm can be easily identified in the framework of the DRM: all languages share the two kinds of

oscillators and hence they are prone to exhibit both tendencies towards stress and syllable timing, although different patterns of syllable-sized durations are found due to differences in coupling.

5. Acknowledgments

This paper is an extended version of a paper presented in *Interspeech 2009* (Barbosa, Viana & Trancoso 2009). The present version added several features such as an introduction explaining the DRM, a thorough discussion on the relation between the model and speech rhythm distinction, deeper explanation of techniques used, and an extended analysis of the data of BP including the ANCOVA. The author acknowledges a grant from CNPq (300371/2008-0). This work was also supported by FCT project PTDC/PLP/72404/2006. INESC-ID Lisboa had support from the POSI Program of the "Quadro Comunitário de Apoio III". Juva Batella is acknowledged for adapting the EP text to BP, and all the speakers are thanked for their time. Wellington da Silva is thanked for the segmentation and labeling of the rest of the corpus of BP.

References

Arantes, P. 2010. Integrando produção e percepção de proeminências secundárias numa abordagem dinâmica do ritmo da fala. PhD diss., State University of Campinas.

Barbosa, P.A. 1996. At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration. In *Proceedings of the 1st ETRW on Speech Production Modeling*. Autrans, France, 85-88.

Barbosa, P.A. 2006. *Incursões em torno do ritmo da fala*. Campinas: Pontes/Fapesp.

Barbosa, P.A. 2007. From syntax to acoustic duration: a dynamical model of speech rhythm production. *Speech Communication* 49: 725–742.

Barbosa, P.A. 2010. Automatic duration-related salience detection in Brazilian Portuguese read and spontaneous speech. In *Proceedings of the Speech Prosody 2010*. 100067**:**1-4. http://www.speechprosody2010.illinois.edu/papers/100067.pdf

Barbosa, P.A., Viana, M.C. & Trancoso, I. 2009. Cross-variety Rhythm Typology in Portuguese. In *Proceedings of Interspeech 2009 - Speech and Intelligence.* London: Causal Productions, 1011-1014.

Boersma, P. & Weenink, D. 2008. Praat: doing phonetics by computer (Version 5.0.35). Computer program. http://www.praat.org

Classe, A. 1939. *The Rhythm of English Prose*. Oxford: Blackwell.

Cowan, N. 1997. *Attention and memory. An integrated framework*. New York: Oxford University Press.

Dogil, G. & Braun, G. 1988. *The PIVOT model of speech parsing*. Wien: Verlag.

Fraisse, P. 1968. Les Rythmes. *Journal Français d'Oto-Rhino-laryngologie* supplément 7: 23-33.

Frota, S., Vigário, M. & Martins, F. 2002. Language Discrimination and Rhythm Classes: evidence from Portuguese. In B. Bel & I. Marlien (eds), *Proceedings of the Speech Prosody 2002 Conference*. Aix-en-Provence, France, 315-318.

Goldman, J.-Ph. 2007. EasyAlign: a Semi-Automatic Phonetic Alignment Tool under Praat. Computer program. http://latcui.unige.ch/phonetique

Kelso, J.A.S. 1995. *Dynamic patterns. The self-organisation of brain and behavior*. Cambridge: MIT Press.

Lehiste, I. 1970. *Suprasegmentals*. Cambridge: MIT Press.

lingWAVES. http://www.wevosys.com/products/lingwaves/lingwaves.html

Massini, G. 1991. A duração no estudo do acento e do ritmo em português. Master thesis, University of Campinas.

O'Dell, M. & Nieminen, T. 1999. Coupled Oscillator Model of Speech Rhythm. In *Proceedings of the XIVth International Conference of Phonetic Sciences*. San Francisco, 1075-1078.

O'Dell, M., Lennes, M. & Nieminen, T. 2008. Hierarchical levels of rhythm in conversational speech. In P.A. Barbosa, S. Madureira & C. Reis (eds), *Proceedings of the Speech Prosody 2008 Conference*. Campinas, Brazil, 355–358.

R project. http://www.r-project.org/

Sauvanet, P. 2000. *Le Rythme et la raison. Rythmologiques.* Paris: Éditions Kimé.

Scott, S.K. 1993. Perceptual centres in speech: an acoustic analysis. PhD diss., University College London.

Valéry, P. 1973. *Cahiers*. Paris: Gallimard. tome 1.

Whitehead, A.N. 1919. *An Enquiry concerning the Principles of Natural Knowledge*. Cambridge: University Press of Cambridge.

DB-IPIC

AN XML DATABASE FOR THE REPRESENTATION OF INFORMATION STRUCTURE IN SPOKEN LANGUAGE

Alessandro Panunzi, Lorenzo Gregori
University of Florence

1. Introduction

1.1 Theoretical framework

In this study we will present a database comprised of a corpus of 74 texts (124735 total words) chosen from the Informal section of Italian C-ORAL-ROM (Cresti & Moneglia 2005; Cresti et al. 2005). The whole corpus has been tagged with respect to the informational structure, and it has been exploited to build a queryable XML database (DB-IPIC) for the study of linear relations among Informational Units in spoken language. The model has also been applied to a subset of C-ORAL-BRASIL corpus (Raso & Mello 2010; Raso & Mello in press), in order to provide statistics for the comparison of informational structure between Italian and Brazilian Portuguese (Mittmann & Raso, in this volume).

The theoretical basis for the database building is the *Language into Act Theory* and the *Informational Patterning Theory* (Cresti 2000, Cresti & Moneglia 2010). Both of these paradigms form a unitary theoretical framework that derives from Austin's *Speech Act Theory* (Austin 1962) and proposes two general hypotheses.

The first one is that spoken language is governed by pragmatic principles (Cresti 1987). Two distinct (but not independent) pragmatic levels operate within the oral performance: a "macro-pragmatic" one, which deals with Speech Act production, and a "micro-pragmatic" one, which deals with the informational structure. The second hypothesis is that the pragmatic features related to these levels are marked and encoded by prosodic phenomena. This regards both the segmentation of the units and their pragmatic values.

At the macro-pragmatic level, the oral performance is structured into Utterances, which correspond to the pragmatic referring unit for spoken language. Utterances

Mello H., Panunzi A., Raso T. (eds), *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation*

are sequences of words that can be pragmatically interpreted, each one corresponding to a Speech Act. On the prosodic side, an Utterance corresponds to a Terminated Sequence (TS), which ends with a perceptually identifiable terminal break.

The definition of the units operating within the micro-pragmatic level is strictly connected with the general assumptions made at the macro-pragmatic one. The informational patterning deals in fact with the features and the modalities of the Speech Act performance: the core Informational Unit (IU) of an Utterance, called Comment, corresponds indeed to the expression of an illocutionary force.

Since the Comment carries the information that ensures the interpretability of a speech sequence, its presence is the necessary and sufficient for the performing of an Utterance. In other words, an Utterance can be costituted by a single Comment. Even if other optional IUs take place in the Utterance, the Comment is the only one that cannot be erased without compromising the interpretation of the whole sequence.

The optional IUs can be divided into two main classes: the textual units, that participate in the construction of the semantic content of the Utterance (Topic, Appendix, Parenthesis, Introducer), and the dialogical units, that are devoted to the successful pragmatic performance of the Utterance in the communicative context (Incipit, Phatic, Allocutive, Conative, Connector; the complete tagset for the IUs is given at paragraphs 2.2 and 2.3, with definitions).

The identification of IUs depends on the internal prosodic parsing of the Utterance into Tone Units (TUs), which are perceptually recognizable through the presence of a non-terminal break. In this respect, the sequence of TUs creates a prosodic pattern, i.e. a model that combines different units in a linear structure, following a unitary programming; the prosodic pattern tendentially corresponds to the informational pattern that gives structure to the Utterance. From an informational point of view, an Utterance corresponds, then, to an informational pattern. Two main cases can be distinguished:

- simple Utterances, whose informational patterns contain only the Comment IU;
- compound Utterances, whose informational patterns contain also optional IUs.

The role of the prosody in the encoding of the pragmatic features is not limited to the parsing of the units, but it also extends to the marking of specific values in both the macro- and the micro- levels. Each language conventionally encodes various illocutionary types (e.g. assertion, question, order, suggestion) by means of dedicated prosodic profiles. In this sense, the prosodic form of the Comment within an Utterance is the formal mark of its specific illocutionary force. Moreover, the prosodic profiles of the optional IUs identify their informational value through their specific and differential forms.

1.2 Application to corpus analysis

Two main principles emerge from the adopted analysis framework, each one dealing with the relationship between pragmatic and prosody:

- the *illocutionary principle*: each Utterance expresses an illocutionary value and corresponds to a prosodically TS;
- the *informational patterning principle*: each Utterance consists of a pattern of IUs that is roughly isomorphic to a pattern of TUs (see paragraph 2.3 , and in particular Table 4, for further details and exceptions).

These principles state that it is possible to carry out corpus-based studies regarding pragmatic features of spoken language starting from the positive perceptual data given by the prosody (Scarano 2009; Moneglia 2011).

As a matter of fact, the segmentation of the speech flow into discrete events is one of the main problems for the analysis of spoken resources. An operative definition of the reference pragmatic units (e.g in terms of Speech Act units) is far from being widely agreed, and their direct identification within oral corpora remains a strongly underdetermined task. On the contrary, prosodic breaks are clearly perceived by speakers. As shown in previous works (Moneglia et al. 2005), their identification has a fairly high degree of inter-annotator agreement (around 95%).

In our framework, the intonational grouping correlates with pragmatical features. The perception-based prosodic tagging can be then used as a heuristic, in order to positively identify the reference pragmatic units of the spoken language: terminal breaks delimit the Utterances, while non-terminal breaks delimit the IUs (Moneglia 2005). Therefore, the informational analysis starts with the prosodic identification of a TS and the related TUs within the speech flow, by means of a perceptual judgment. On this basis, the *root* TU, which contains the necessary and sufficient information for the interpretation of the TS, is identified as Comment, and after this the annotator can assing an informational value to the other TUs.

For the building of the DB-IPIC, the workflow proceeded through four main stages, which will be described in detail in the following paragraph:

- the session recording;
- the session transcription and the annotation of prosodic boundaries (both terminal and non terminal);
- the text-to-speech alignment;
- the informational tagging of each TU;
- the data conversion in XML format.

2. The tagging procedure and the database building

2.1 Prosodic parsing

The prosodic parsing is performed during the transcription task. Its primary objective is to determine the reference units of spoken language by means of the identification of tonal breaks, which are variations in the speech continuum such as to cause its parsing into discrete units (Moneglia 2005: 17).

The transcripion is performed using an adaptation of the CHAT format (MacWhinney 2000, Moneglia & Cresti 1997). Different kinds of terminal breaks are reported:

- the question mark (?) is used to delimit a TS with a clear interrogative prosodic profile (1);
- suspension points (…) delimit a TS voluntarily interrupted by the speaker, who performs a suspensive prosodic profile (2);
- the plus sign (+) is used for unintentionally interrupted TSs (e.g. interrupted by the interlocutor); in this case, the speaker program is broken and the interpretability of the sequence can be compromised (3);
- the double slash (//) is the main tag for terminal breaks, and marks all TSs that do not belong to the previous classes (4);

(1) *SMN: e che lavoro fai? (ifamdl06, 2)
[*SMN: and what is your job?]

(2) *IDA: ma si sono scambiati i numeri di telefono… (ifamdl20, 181)
[*IDA: but they have exchanged their phone numbers…]

(3) *MAX: volevo sapere + (ifamcv27, 23)
[*MAX: I'd like to know +]

(4) *LUC: questo lo puoi fare anche il giorno prima // (ifammn11, 129)
[*LUC: you can even do this the day before //]

Standard non-terminal breaks are marked by a single slash (/), and delimit TUs.

(5) *VAL: secondo me / lui è chiarissimo / a lezione // (ifamcv27, 365)
[*VAL: in my opinion / he is very clear / during the lesson //]

Retracting phenomena (i.e. false starts) are also marked in the transcripts through the [/*n*] symbol, where *n* corresponds to the number of retracted words. Retracting

marks can be considered as non-terminal breaks. Anyway, since the word sequences involved in false starts are "discarded" by the speaker, they do not contribute to the informational patterning and to the semantic content of the Utterance (see Table 4 below, Interrupted units).

(6) *MAR: è un gioco &diffi [/4] è un gioco da grandi // (ifamcv09, 275)
[*MAR: this is a &diffi [/4] this is a game for adults //]

The prosodically annotated transcripts are then aligned according to the terminal breaks. The alignment procedure is performed using the WinPitch software (Martin 2005), and allows the simultaneous access of both textual data and sound.

2.2 Informational annotation: types of Comment and reference units

After the prosodic parsing, each TU is tagged with its own informational value. This procedure starts with the identification of the Comment unit. With respect to this, different structures can be identified within a TS. Usually, in a TS there is only one Comment IU that bears the illocutionary force of the Utterance. However, it is also possible that more than one IU carrying an illocutionary value is present in a TS. These cases correspond to two different phenomena.

First, a TS can contain a Multiple Comment (CMM), i.e. a compositional unit formed by two or more Comments, each one carrying an illocutionary force, linked together by a conventional prosodic model. In this sense, the Multiple Comment creates a compositional *illocutionary pattern*, i.e. a model, codified by the language, that allows the linking of two illocutionary values, and that produces a meta-illocutionary "rhetoric" effect, such as:

- strengthening (7);
- binding relation (8);
- comparison (9);
- alternative and double directive (10);
- list (11).

(7) *LIA: il cuoco /^{CMM} sì //^{CMM} (ifamcv01, 163)
[*LIA: the cook /^{CMM} yeah //^{CMM}]

(8) *CLA: se son qui /^{CMM} è inutile andare //^{CMM} (ifammn03, 490)
[*CLA: if they're not here /^{CMM} it's useless to go//^{CMM}]

(9) *LUI: nel senso la zona espositiva è da una parte /^{CMM} la zona dei servizi è da un'altra //^{CMM} (ifamcv16, 195)
[*LUI: I mean the exibition area is on one side /^{CMM} the service area is on another side //^{CMM}]

(10) *CIC: le metti /^{CMM} o no?^{CMM} (ifamcv14, 202)
[*CIC: do you put them /^{CMM} or not?^{CMM}]

(11) *DAN: allora mise il cappellino /^{CMM} il cappuccio /^{CMM} e parti //^{CMM} (ifammn25, 15)
[*DAN: then she put on the little cap /^{CMM}the hood /^{CMM} and she left //^{CMM}]

Second, a TS can contain a sequence Comment IUs characterized by a homogeneous and weak illocutionary value. In this case, each IU is considered as a Bound Comment, and does not form a compositional unit with the other ones. In the annotation practice, all the Bound Comments are labeled with the COB tag but the last one, which is labeled as COM.

While a Multiple Comment is a patterned sequence that properly performs a prosodic and informational model, the chain of Bound Comments is not: it is indeed tied up by a progressive adjunction of oral text, out of any informational programming.

(12) *ROS: e mi dà anche delle soddisfazioni /^{COB} perché è un lavoro creativo //^{COM} (ifamdl07, 27)
[*ROS: and it also gives me satisfaction /^{COB} because it is a creative job //^{COM}]

(13) *DAN: il lupo invece prese la via più breve /^{COB} entrò nella casa della nonna /^{COB} la vide /^{COB} e se la mangiò //^{COM} (ifammn25, 33)
[*DAN: on the contrary the wolf took the shortest way /^{COB} went into the grandmother's house /^{COB} he saw her /^{COB} and he ate her //^{COM}]

The following table contains the definitions and the labels for Comment, Multiple Comment and Bound Comment units.

Table 1. Informational tagset, first part: Comment units

Name	**Tag**	**Definition**
Comment	COM	Comment IU accomplishes the illocutionary force of the Utterance, and it is therefore necessary and sufficient to perform an utterance
Multiple-Comment	CMM	A complex IU comprised of two or more Comments, forming an *illocutionary pattern*
Bound Comment	COB	A sequence of Bound Comments with weak illocutionary force, produced by progressive adjunctions following the flow of thought, out of any model of informational patterning

Since the Utterance has been defined as a patterned entity performing a single Speech Act and following a unitary programming, the presence of complex illocutionary structures within a TS gives rise to the need of reconsidering the definition of the spoken language referring units.

For what regards TSs that contain a Multiple Comment, we must consider that their informational structure is properly patterned, and that Multiple Comments perform a Speech Act with a coherent intentionality, as well as simple Comments do. For this reasons, they can be considered as Utterances.

On the contrary, a TS that contains a sequence of Bound Comments cannot be considered as an Utterance, since it is not structured as an informational pattern and it carries a weak illocutionary value. In this case, a different reference unit has been introduced in the theoretical framework: the Stanza (cf. Cresti 2000; Panunzi & Scarano 2009; Cresti 2009), which corresponds to a linguistic "activity" whose primary intention is the production of an oral text (while the primary intention of an Utterance is to perform a Speech Act).

In brief, from an informational point of view, a TS may correspond to different referring units, and namely:

- an Utterance, if it contains a simple Comment or a Multiple Comment, it is prosodically and informationally patterned and it is aimed at the performing of a Speech Act;
- a Stanza, if it contains a sequence of Bound Comments, it is not prosodically and informationally patterned and it is aimed at the production of an oral text.

2.3 Informational annotation: optional units

After the identification of the necessary units in a TS, the tagging procedure takes into account all the other TUs in order to provide them with an informational label. Table 2 and Table 3 introduce the tagset for, respectively, the optional textual units and the dialogical ones. After each table, examples for the various IUs are reported.

Table 2. Informational tagset, second part: Textual units

Name	Tag	Definition
Topic	TOP	It identifies the domain of application for the illocutionary act expressed by the comment, providing the Speech Act with a cognitive reference and allowing the Utterance displacement from the actual context
Topic List	TPL	A chain of Topics forming a pattern of Topics
Appendix of Comment	APC	It integrates the text of the Comment and concludes the Utterance
Appendix of Topic	APT	It gives a delayed integration of the information given in the Topic adding specification for the addressee
Parenthesis	PAR	It adds information to the utterance with a meta-linguistic value having "backward" or "forward" scope; always bears a modal value.
Locutive Introducer	INT	It is used for introducing a sequence of IUs that have a strong and unitary "point of view", as in reported speech and reported thought

(14) *ANN: a Firenze /TOP c'hanno tutti queste idee …COM (ifamcv26, 160)
[*ANN: in Florence /^{TOP} everybody has these ideas …^{COM}]

(15) *CLA: quando arrivano su al villaggio /^{TPL} gnudi /^{TPL} quest' omini /^{TPL} sono una bellezza incredibile //^{COM} (ifammn03, 517)
[*CLA: when they arrive to the village /^{TPL} naked /^{TPL} these man /^{TPL} they are amazingly beautiful //^{COM}]

(16) *MAX: quand'è stata fatta /^{COM} questa qui ?^{APC} (ifamcv01, 88)
[*MAX: when has it been done /^{COM} this thing ?^{APC}]

(17) *MIC: ma gli accessi principali /^{TOP} in questa zona qua /^{APT} quali sarebbero ?^{COM} (ifamcv16, 133)
[*MIC: but the main entries /^{TOP} in this area /^{APT} what would they be ?^{COM}]

(18) *MAR: m'ha richiamato /$^{i\text{-}COM}$ invece /^{PAR} dopo una settimana /^{COM} credo //^{PAR} (ifammn23, 156)
[*MAR: he called me again /$^{i\text{-}COM}$ on the contrary /^{PAR} after a week /^{COM} I believe //^{PAR}]

(19) *LIA: dice /^{INT} guarda come li spolvera //$^{COM\_r}$ (ifamcv01, 775)
[*LIA: he says /^{INT} look how she dusts them //$^{COM\_r}$]

Table 3. Informational tagset, third part: Dialogic units

Name	Tag	Definition
Incipit	INP	It opens the communicative channel bearing a contrastive value starting a dialogic turn or an utterance
Conative	CNT	It pushes the listener to take part in the dialogue in an adequate way, or stops his non collaborative behavior
Phatic	PHA	It is dedicated to controlling the communicative channel, ensuring its maintenance; it stimulates the listener to the social cohesion needed by the dialogical exchange and/or ensures the reception of the utterance
Allocutive	ALL	It specifies to whom the message is directed keeping his attention. Simultaneously it plays a cohesive and empathic function, bringing the interlocutor to share the point of view of the utterance
Expressive	EXP	It works as an emotional support. It stresses the sharing of a common social affiliation with the interlocutor, searching for social cohesion.
Discourse Connector	DCT	It zips different parts of the discourse (e.g. utterances within a turn), signaling to the addressee that the discourse is going on and that the entity which follows holds a relation with the previous ones.

(20) *SMN: quindi /INP ami molto gli animali //COM (ifamdl06, 98)
[*SMN: so /^{INP} you love animals so much //^{COM}]

(21) *LIA: qui /TOP eravamo a Venezia /COM guarda //CNT (ifamcv01, 919)
[*LIA: here /^{TOP} we were in Venice /^{COM} look //^{CNT}]

(22) *CLA: e non era facile /^{COM} sai //^{PHA} (ifammn02, 302)
[*CLA: and it wasn't easy /^{COM} you know //^{PHA}]

(23) *GIO: Giulia /^{ALL} non urlare //^{COM} (ifamcv24, 205)
[*GIO: Giulia /^{ALL} don't scream //^{COM}]

(24) *ALE: mannaggia /^{EXP} ora come si fa?^{COM} (ifamcv15, 60)
[*ALE: damn it /^{EXP} what can we do now?^{COM}]

(25) *SIM: inoltre /^{DCT} mi dovresti togliere una curiosità //^{COM} (ifamcv07, 76)
[*SIM: moreover /^{DCT} you should satisfy my curiosity //^{COM}]

The last part of the tagset comprehends the TUs that do not have an informational value. Among them, the most prominent case deals with the possibility that an IU is scanned in two (or even more) TUs. This is mostly due to performance reasons: for instance, an IU with a "heavy" locutive content may require two TUs to be performed. In this case, the prosodic pattern and the informational one are not strictly isomorphic. The informational tagging conventionally considers the TUs on

the right as "scanning" units (SCA), and marks the informational value only for the last one (26).

Other cases of non-informational labeling for TUs are Interrupted units, Time Taking and Unclassified units, as shown in Table 4.

Table 4. Informational tagset, fourth part: non-informative units

Name	Tag	Definition
Scanning	SCA	It occurs when the corresponding prosodic unit has no informational function and its locutive content is part of a larger IU (by default occurring on its right)
Interrupted	EMP	Interrupted units which cannot be evaluated
Time Taking	TMT	Time taking units for programming needs
Unclassified	UNC	Unclassified Units

(26) *GIU: il prete lo chiamava /^{SCA} sempre a spazzare la chiesa //^{COM} (ifamcv20, 24)
[*MAR: the priest called him /^{SCA} always to sweep the church //^{COM}]

(27) *ELA: vicino a +^{EMP} (ifamcv01, 52)
[*ELA: near to +^{EMP}]

(28) *PRE: e essenzialmente /^{TOP} &he /^{TMT} la modifica riguarda due aree //^{COM} (ipubcv04, 44)
[*PRE: and basically /^{TOP} &he /^{TMT} the modification regards two areas //^{COM}]

2.4 DB building

After the informational tagging procedure, all transcripts are automatically PoS-tagged through the TreeTagger software and then converted to the XML format, following a schema that has been specifically developed for the DB.

The choice of the XML format is motivated by several reasons. First, the XML format allows an efficient standardization of the annotated data and a formal validation. Moreover, XML is able to encode information that requires different kinds of representation (category, structural and relational information) and its elements are organized into a hierarchic model, which adequately fits with the representation of different levels of our analysis. Finally, the XML "family" comprehends query languages directly applicable to the annotated texts.

For each recording session, an XML document has been created, comprehending both recording metadata and the annotated transcript. The XML schema adopted for the representation of the DB is structured as follows.

At the lower layer there are the tokens, which comprehend the following elements:

- <word> for each word form, with “pos” and “lemma” attributes that derive from the PoS tagging;
- <frag> for fragmented words;
- <paralinguistic> for non-linguistic elements that occur within the speech flow, such as laughs, grumbles, coughs etc.;
- <break> for prosodic breaks (the “type” attribute specifies wheter the break is terminal, non terminal or a retracting break);
- <notation> for all the other symbols used for transcription, such as overlaps and pauses.

The further layers of annotation represent the prosodic groupings in a hierarchical structure, which is organized in three levels:

- <tone_unit> groups a sequence of tokens (the informational value of the unit is identified in the “inf” attribute);
- <term_seq> groups one or more Tone Units within a prosodically terminated sequence (the “type” attribute specifies wheter the terminated sequence corresponds to an Utterance or to a Stanza);
- <turn> groups an uninterrupted series of Terminated Sequences uttered by a single speaker.

A sample of an XML document with all the annotation levels for a single turn (one utterance divided into two prosodic/information units) follows:

```
<turn speak="EDO">
        <term_seq num="1" type="utt">
          <tone_unit inf="COM">
                    <word lemma="guardare" pos="VER:fin">guarda</word>
                    <word lemma="chi" pos="WH">chi</word>
                    <word lemma="ci" pos="ADV">c'</word>
                    <word lemma="essere" pos="VER:fin">è</word>
                    <break type="nonterminal">/</break>
          </tone_unit>
          <tone_unit inf="ALL">
                    <word lemma="nonna" pos="NOUN">nonna</word>
                    <break type="terminal">//</break>
          </tone_unit>
        </term_seq>
</turn>
```

All the annotated transcripts in XML format have been inserted in a database. The resource runs on the eXist engine, an open source database management system that stores data according to the XML data model and features index-based XPath/XQuery processing.

A user-friendly web interface has been developed to allow the extraction of informational patterns (Gregori 2011). The interface also allows the user to filter data with respect to session metadata (Figure 1).

Figure 1. Query interface

The results of queries via the web interface are shown in the CHAT format (Figure 2). Audio is directly accessible, through the exploitation of the alignment data.

The following paragraph introduces some general data extracted from the database via the query interface, mainly focusing on the pragmatic referring units of spoken language: Utterances and Stanzas.

Figure 2. Query results

3. Data from the corpus

3.1 Data on the main referring units: Utterances vs. Stanzas

The general data regarding the database size, with respect to the main tagging elements, are shown in Table 5. The corpus consists of about 2/3 dialogic interactions (dialogues between two interlocutors and conversations among three or more interlocutors), and 1/3 of monologic ones.

Table 5. Corpus size per unit

	sessions	turns	TSs	TUs	words
dialogic	47	8823	15742	31081	78394
monologic	27	924	5265	16777	46341
TOTAL	74	9747	21007	47858	124735

Starting with these data, we will focus on the TSs. The first observation that can be made is that a very relevant number of TSs are interrupted (2889, which corresponds to 13.7% of the total). The estimation of this percentage has made with all the TSs that lack any Comment IU and therefore do not perform a Speech Act. Since these

TSs are not interpretable, they were excluded from further estimations. The completed TSs in our corpus are then 18118.

Given this, the first measure about the structuring of TSs takes into account the differentiation between Utterances and Stanzas. Table 6 reports these data, distinguishing between dialogical interactions and monologic ones. The percentages reported in the table refer to the total constituted by a single row (e.g. the first row reports the percentages of Utterances vs. Stanzas within the sole dialogic interactions).

Table 6. Utterances and Stanzas

	Utterances	**%**	**Stanzas**	**%**
dialogic	12694	94.1%	791	5.9%
monologic	3779	81.6%	854	18.4%
TOTAL	16473	90.9%	1645	9.1%

It emerges from the data that the number of Stanzas in monologic interactions is much higher than the one for dialogues and conversations (more than 3 times higher). These data reflect the fact that the text construction in monologues is more structured than in dialogic interactions: since Stanzas are devoted to the production of an oral text, they are much more frequent in the contexts where text construction is more relevant, as in monologues.

3.2 Utterances with simple Comment and Multiple Comment

If we consider only the Utterances, it is possible to observe data about the distinction between those with single Comments (henceforth COM-Utterances) and those with Multiple Comments (henceforth CMM-Utterances), as reported in Table 7.

Table 7. COM-Utterances and CMM-Utterances

	COM-Utt	**%**	**CMM-Utt**	**%**
dialogic	11438	90.1%	1258	9.9%
monologic	3413	90.4%	364	9.6%
TOTAL	14851	90.2%	1622	9.8%

The data show that the distribution of COM-Utterance and CMM-Utterance remains constant with respect to the variation between dialogic and monologic interaction.

If we consider also the distinction between Simple Utterances and Compound Utterances, other interesting data emerge. For this computation, we considered as Simple Utterances the ones composed by the only Comment IU (or Multiple Comment) and other non-informational units (SCA, EMP, TMT). The following examples show the types distinguished in the Table 8: simple COM-Utterance (29),

compound COM-Utterance (30), simple CMM-Utterance (31) and compound CMM-Utterance (32):

(29) *ZIA: te lo dico dopo //^{COM} (ifammn01, 106)
[*ZIA: I will tell you about it later //^{COM}]

(30) *MIC: ma un filo d'acqua /^{TOP} dove /^{COM} scusa ?^{PHA} (ifamcv16, 54)
[*MIC: but a trickle of water /^{TOP} where /^{COM} sorry ?^{PHA}]

(31) *ALE: lei è una biondina /^{CMM} lui con gli occhi azzurri //^{CMM} (ifamcv15, 313)
[*ALE: she is a fair-haired girl /^{CMM} he has blue eyes //^{CMM}]

(32) *ANT: che pensi /^{ANT} questo qui /^{TOP} lo faceva bene /^{CMM} o lo faceva male //^{CMM} (ifamdl01, 502)
[*ANT: what do you think /^{ANT} this one /^{TOP} he did it well /^{CMM} or he did it wrong //^{CMM}]

Table 8. Simple and compound Utterances

	Simple	**%**	**Compound**	**%**
COM-Utt	9927	66.8%	4924	33.2%
CMM-Utt	1017	62.7%	605	37.3%
TOTAL	10944	66.4%	5529	33.6%

Also the percentages regarding the informational complexity are similar between COM-Utterances and CMM-Utterances. Again, Stanzas show very different values: just the 30% of them is formed by only Bound Comments units, while 70% contains also an optional textual or dialogic IU.

The last two sets of data regard the distribution of the Textual units within COM-Utterances, CMM-Utterances and Stanzas. Table 9 and Table 10 show the numbers and the percentages of referring units in which, respectively, the different textual IUs and dialogic IUs occur.

Table 9. Presence of textual IUs within Utterances and Stanzas

IU	**COM-Utt**	**%**	**CMM-Utt**	**%**	**Stanza**	**%**
TOP	2046	13,78%	236	14,55%	539	32,77%
TPL	90	0,61%	7	0,43%	21	1,28%
APC	735	4,95%	70	4,32%	102	6,20%
APT	102	0,69%	4	0,25%	23	1,40%
PAR	678	4,57%	93	5,73%	312	18,97%
INT	430	2,90%	140	8,63%	220	13,37%

Table 10. Presence of dialogic IUs within Utterances and Stanzas

IU	COM-Utt	%	CMM-Utt	%	Stanza	%
PHA	1384	9,32%	132	8,14%	345	20,97%
ALL	161	1,08%	13	0,80%	12	0,73%
INP	893	6,01%	98	6,04%	218	13,25%
CNT	191	1,29%	53	3,27%	27	1,64%
EXP	103	0,69%	11	0,68%	13	0,79%
DCT	224	1,51%	54	3,33%	223	13,56%

In the majority of cases the percentages are similar between COM and CMM-Utterances, while Stanzas record a higher number of optional IUs. This is true in particular for the most frequent textual units (Topic, Parenthesis and Locutive Introducer) and dialogic ones (Phatic, Incipit and Dialogic Connectors).

3.3 Final remarks

The whole set of extracted data allow us to sketch an overall distinction between Utterances (both COM- and CMM- ones) and Stanzas, following quantitative parameters:

- COM-Utterances and COB-Utterances are similar for what regards their distribution within dialogic interactions and monologic ones, while Stanzas are 3 times more frequent in monologues;
- COM-Utterances and COB-Utterances are similar for what regards the measure of their complexity: around 65% of them are simple and 35% are compound; on the contrary, 70% of Stanzas have a complex structure and contain at least an optional IU;
- COM-Utterances and COB-Utterances show similar percentages for what regards the presence of optional IUs, while Stanzas contain them more frequently.

These results give a quantitative consistency to the distinction between two pragmatic referring units, Utterance and Stanza, and they constitute an *a posteriori* validation of the criteria adopted for their distinction.

References

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

Cresti, E. 1987. L'articolazione dell'informazione nel parlato. In *Gli Italiani Parlati: Sondaggi sopra la lingua d'oggi.* Firenze: Accademia della Crusca, 27-90.

Cresti, E. 2000. *Corpus di italiano parlato*, 2 voll., CD-ROM. Firenze: Accademia della Crusca.

Cresti, E. 2009. Unità di analisi testuale e caratteri costruttivi nell'italiano parlato (spontaneo) e scritto (letterario). Ricerche corpus-based. In A. Ferrari (ed.), *Sintassi storica e sincronica dell'italiano. Subordinazione, coordinazione, giustapposizione*, vol. 2. Firenze: Cesati, 713-732.

Cresti, E. & Moneglia, M. (eds) 2005. *C-ORAL-ROM. Integrated reference corpora for spoken romance languages*, DVD + vol. Amsterdam: Benjamins.

Cresti, E. & Moneglia, M. 2010. Informational patterning theory and the corpus-based description of spoken language. The compositionality issue in the topic-comment pattern. In M. Moneglia & A. Panunzi (eds), *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*. Firenze: FUP.

Cresti, E., Panunzi, A. & Scarano, A. 2005. The Italian corpus. In E. Cresti & M. Moneglia (eds), *C-ORAL-ROM: integrated reference corpora for Spoken Romance Languages*. Amsterdam/Philadelphia: John Benjamins, 71-110.

eXist. http://exist.sourceforge.net/

Gregori, L. 2011. Database XML per annotazione multilivello del corpus di parlato spontaneo LABLITA. MA thesis, Università degli Studi di Firenze.

MacWhinney, B. 2000. *The CHILDES Project*: *Tools for Analyzing Talk.* 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Martin, Ph. 2005. WinPitch Corpus: a text-to-speech analysis ans alignment tool. In E. Cresti & M. Moneglia (eds), *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: John Benjamins, 40-51.

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/Philadelphia: John Benjamins, 1-70.

Moneglia, M. 2011. Spoken corpora and pragmatics. In H. Mello & S. Gries (eds), *Brasilian Journal of applied linguistics, Revista brasileira de linguìstica aplicada* vol. 11, n. 2, 479-519.

Moneglia, M. & Cresti, E. 1997. Intonazione e criteri di trascrizione del parlato. In U. Bortolini & E. Pizzuto (eds), *Il progetto CHILDES Italia*, vol. II. Pisa: Il Cerro, 59-90.

Moneglia, M., Fabbri, M., Quazza, S., Panizza, A., Danieli, M., Garrido, J.M. & Swerts, M. 2005. Evaluation of consensus on the annotation of terminal and non-terminal prosodic breaks in the C-ORAL-ROM Corpus. In E. Cresti & M. Moneglia (eds), *C-ORAL-ROM: integrated reference corpora for Spoken Romance Languages*. Amsterdam/Philadelphia: John Benjamins, 257-276.

Panunzi, A. & Scarano, A. 2009. Parlato spontaneo e testo: analisi del racconto di vita. In L. Amenta & G. Paternostro (eds), *I parlanti e le loro storie. Competenze linguistiche, strategie comunicative, livelli di analisi: Atti del Convegno Carini-Valderice, 23-25 October 2008*. Palermo: Centro di studi filologici e linguistici siciliani, 121-132.

Raso, T. & Mello, H. 2010. The C-ORAL-BRASIL corpus. In M. Moneglia & A. Panunzi (eds), *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*. Università degli studi di Firenze, 193-213.

Raso, T. & Mello, H. in press. *C-ORAL-BRASIL I. Corpus de referência para a fala espontânea informal do português do Brasil*. Belo Horizonte: UFMG.

Scarano, A. 2009. A The prosodic annotation of C-ORAL-ROM and the structure of information in spoken language. In L. Mereu (ed.), *Information structures and its interfaces*. Berlin and New York: Mouton de Gruyter, 51-74.
Treetagger. http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/
WinPitch. http://www.winpitch.com/
XML. http://www.w3.org/XML/

THE C-ORAL-BRASIL INFORMATIONALLY TAGGED MINI-CORPUS[1]

Maryualê M. Mittmann, Tommaso Raso
Universidade Federal de Minas Gerais (Brazil)

1. Introduction

This paper has two main goals:

- To present a corpus of small proportions that constitutes a sample extracted from the C-ORAL-BRASIL corpus (Raso & Mello 2010; Raso & Mello in press) for spontaneous spoken Brazilian Portuguese. This corpus was tagged with respect to the informational structure following the Language into Act Theory (Cresti 2000) and therefore allows some first consideration about the information structure of Brazilian Portuguese. A comparable mini-corpus was selected for Italian from the Italian C-ORAL-ROM. In the paper we give some first results of the comparation of the two mini-corpora.
- To discuss some interesting aspects of the prosodic annotation of the C-ORAL-BRASIL corpus observing the corrections of the annotation done during the information tagging. The informational tagging is a different perspective from that of the prosodic annotation, and the study of the corrections of the prosodic annotation during the process of informational tagging is useful for better understanding both the perceptual aspects of the prosodic annotation and the cognitive aspects of the informational tagging.

The Brazilian sample is referred as Brazilian mini-corpus, and is a 15 percent (in number of words) portion of C-ORAL-BRASIL informal section[2]. The Italian

[1] This research was financed by CNPq and Fapemig.

Mello H., Panunzi A., Raso T. (eds), *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation*

sample (Italian mini-corpus) was extracted from the C-ORAL-ROM Italian corpus (Cresti & Moneglia 2005; Cresti, Panunzi & Scarano 2005), and represent a larger part of the Italian informal corpus.

C-ORAL-BRASIL is a corpus of spontaneous speech of Brazilian Portuguese, coordinated by Tommaso Raso and Heliana Mello. The project is part of an international cooperation and constitutes the fifth branch of the European C-ORAL-ROM project (Cresti & Moneglia 2005). The architecture of the Brazilian corpus follows the same guidelines of the European corpora represented in C-ORAL-ROM, which ensures the comparability of both language resources. The informal section of C-ORAL-BRASIL comprises 139 texts and a total of 208,130 words in 21:08:00 of recording sessions, with a total of 34,167 terminated linguistic sequences (utterances). The informal portion is divided according to the context of the interactions: family/private (105 texts and 159,364 words) and public (34 texts and 48,766 words). Each of these sections is further equally subdivided according to the type of interaction: monologues, dialogues or conversations. Each subsection contains 1/3 of the texts. The diatopic variety represented in C-ORAL-BRASIL is the one of Minas Gerais state, in particular the metropolitan area of its capital Belo Horizonte[3].

The main goal of both the C-ORAL-ROM and the C-ORAL-BRASIL corpora is the documentation of the diaphasic variation, necessary to represent really spontaneous speech. Therefore, besides the variation between private/familiar and public contexts and among the three interactional typologies (monologues, dialogues and conversations), the corpora try to document the largest variation in terms of different interaction situations, so allowing a great variation of activity and, as a consequence, of different speech acts and information structures.

As in C-ORAL-ROM corpora, C-ORAL-BRASIL transcriptions incorporate the annotation of prosodic boundaries proposed by Moneglia & Cresti (1997). The annotation scheme segments the speech flow in two distinct levels. The first level deals with the demarcation of the fundamental entity in spontaneous spoken communication, that is the utterance. The utterance is signaled by a prosodic boundary that bears a conclusive value (terminal prosodic break) and conveys a speech act. The second level refers to the internal structure of the utterance, that can be built by one single tone unit (simple utterance) or by several tone units (compound utterance). Tone units within an utterance are prosodically signaled by boundaries with non-conclusive value (non-terminal prosodic break) (Moneglia & Cresti 1997; Moneglia & Cresti 2006).

---

[2] C-ORAL-BRASIL corpus will contemplate two major sections: one for informal speech and one for formal speech. The informal section is completed and the formal section is in compiling phase.

[3] More detailed information can be found at Raso & Mello (2010) and in press.

The Brazilian and European corpora have been designed to allow the study of illocutions and information structure of spontaneous speech. In order to allow the latter, the Brazilian mini-corpus received a tagging (complementary to the annotation of prosodic segmentation) that associates information functions to each one of the segmented prosodic units (Cresti 1987; Cresti 2000; Cresti & Moneglia 2010). The informational tagging is based in the model proposed by Language into Act Theory (Cresti 1987; Cresti 2000). This model was first implemented in the LABLITA corpus of Spontaneous Spoken Italian (Cresti 2006), from which the Italian corpus is derived.

The process and criteria of compiling the Brazilian mini-corpus is showed in section 2. In section 3 we present the methodology and tagset employed for the information structure annotation. Section 4 features some structural and informational characteristics of spontaneous spoken Brazilian Portuguese derived from the Brazilian mini-corpus. In section 5 we compare some of these results with the Italian mini-corpus. In section 6 the relationship between prosodic and informational annotation is discussed.

2. Strategy and criteria for compiling the Brazilian mini-corpus

In order to study the information structure, we need a corpus that identifies the informational functions of each prosodic unit; in other words, we must have an informationally tagged corpus. Unlike the tagging of part-of-speech, for which there are already many automatic tools, the tagging of information units is done manually. The information tagging of all C-ORAL-BRASIL texts, which comprises more than 61,000 information units, requires a considerable amount of time and human resources. For this reason, in a first stage, we selected a sample of the informal section of the C-ORAL-BRASIL corpus to receive informational tagging, thus enabling studies of informational nature.

The selection of texts followed criteria adopted to ensure a high quality database to perform information structure studies, but at the same time preserving the same basic structure of the entire corpus, so that the results obtained with the mini-corpus could be extrapolated to the whole corpus. Given the impossibility of balancing all the corpus variations in the mini-corpus, the parameters chosen as guidelines to achieve the best possible sample are the following (Raso & Mello 2009):

- Representativeness of typological branch. Dialogues and conversations should be 2/3 of the mini-corpus and monologues should be 1/3. The texts should be good exemplars of the context and text typologies: familiar/private and public dialogues, monologues and conversations.

- Highest possible range of communicative situations and activities. That means that speakers in different texts should perform different tasks, to ensure diaphasic variation.
- High acoustic quality. The quality is determined based on the absence (total or partial) of background noise, no feedback signal, voice clarity, good audio gain and low percentage of overlapping. The calculation of F0 curve must be (almost) always possible.
- Diversity of speakers. The goal is to have a balanced number of male and female voices and, if possible, also ages and school levels.
- Interesting text content. Texts with interesting content lead to higher attention of transcribers. Also, texts with interesting content increase the degree of informativeness within the sample.

The construction of the Brazilian mini-corpus involved the following steps.

1. Session recording, with the participants' consent, in digital format (wav).
2. Text transcription in CHAT format (MacWhinney 2000) with concomitant annotation of prosodic boundaries (Moneglia & Cresti 1997).
3. Review of the transcriptions, that includes the check for the appropriate application of the set transcription criteria (Mello & Raso 2009) and accurate annotation of prosodic breaks, always performed by a person other than the one who did the original transcription.
4. Text-to-spech alignment through software WinPitch (Martin 2005). Each audio file is aligned with the text according to the linguistic sequences marked by terminal prosodic boundary.
5. Informational tagging, performed on the aligned transcripts. During this phase, errors in the transcription and in the annotation of prosodic boundaries were also checked and corrected.
6. Two revisions of informational tagging and further correction of the transcripts.

The annotation of prosodic boundaries was validated in two occasions, once before the beginning of the transcription work and another when all transcriptions and the first revision were completed, but before the further revisions and the informational tagging. The final result of the validation reached a Kappa score agreement (Fleiss 1971) of 0,86, 0,87 for terminal breaks and of 0,78 for non terminal breaks (Raso & Mittmann 2009; Mello et al. in press).

Table 1 presents the information about each text of the Brazilian mini-corpus, indicating the text identification, the communicative situation, the number of male and female participants and duration of the audio file. The monologic group consists of narratives, descriptions and explanations. Monologues are highly elaborated texts, thus featuring less, but more complex, linguistic entities (utterances, illocutionary

patterns and stanzas). Instead, conversations and dialogues comprise texts in which the speech is highly situated and entrenched in the immediate extra-linguistic context, and consequently they feature more utterances, with a less complex structure but with much more speech acts variation. As conversations are concerned, the first two represent the very common situation of friends just chatting.

Text names are composed by terms that indicate: language, context and text type. Thus we have ‘b’ for the Brazilian Portuguese, ‘fam’ to the family/private and ‘pub’ for public context, ‘cv’ for conversation, ‘dl’ for dialogue and ‘mn’ to monologue. Each text receives a double-digit sequential number that identifies it within the section to which it belongs.

Table 1. Situations recorded, number of male and female speakers and duration of texts

Text	**Situation**	**M**	**F**	**Duration**
Total		28	27	03:58:36
Conversations		15	9	01:07:28
bfamcv01	Chat between young friends	4	0	00:07:00
bfamcv02	Chat between elderly ladies	0	3	00:07:51
bfamcv03	Friends play snooker	5	0	00:06:50
bfamcv04	Friends play Pictionary	2	2	00:07:30
bpubcv01	Employees at a blood bank explain their work	1	3	00:08:30
bpubcv02	Political meeting	3	1	00:29:47
Dialogues		6	8	01:45:28
bfamdl01	Two friends do the groceries	0	2	00:14:39
bfamdl02	Two friends pack the recording equipment	1	1	00:07:26
bfamdl03	Couple takes a car trip	1	1	00:10:30
bfamdl04	Maids do the dishes	0	2	00:19:32
bfamdl05	Broker shows apartment to his sister*	1	1	00:11:28
bpubdl01	Engineer and construction worker at construction site	2	0	00:26:08
bpubdl02	Customer and salesman in a shoe store*	1	1	00:15:45
Monologues		7	10	01:05:40
bfammn01	Man tells an alleged true story about a snake	2	0	00:05:02
bfammn02	Grandmother tells grandson stories about her famous uncle	1	1	00:07:23
bfammn03	Father tells family two entertaining stories*	3	3	00:07:08
bfammn04	Woman tells about her experience in the hospital*	0	1	00:06:57
bfammn05	Woman shares the story about her daughter's adoption*	0	2	00:09:52
bfammn06	Man explains its professional trajectory	1	1	00:10:02
bpubmn01	Teacher evaluates her work at public school	0	2	00:19:16

* minor third party interventions.

The speakers' characteristics are almost perfectly balanced. In Table 1 are included also speakers that participate of the situation only for a few moments or that represent the interlocutors of the monologants. But if we consider the main speakers only, the balancing in term of uttered words is much better. The Brazilian mini-corpus features 23 speakers in conversation (one of them appears twice), 14 in dialogues and 7 in monologues. As far as gender is concerned, 25 are males and 19 are females, but the balancing in terms of words is almost perfect, since in conversations, where the number of words for each speaker in considerably smaller than in dialogues and specially in monologues, we have 16 males and only 7 females. The number of females is higher in dialogues (8 versus 6) and in monologues (4 versus 3). Age and school level are also balanced.

For age, we have in conversations 9 A speakers (from 18 to 25 years old), 9 B speakers (from 26 to 40 years old), 4 C speakers (from 41 to 60 years old) and 2 D speakers (more than 60 years old); in dialogues 4 A speakers, 3 B speakers, 6 C speakers and 1 D speaker; in monologues, 4 C speakers, 2 D speakers and 1 B speaker. For school level, speakers are divided in three different levels: level 1 refers to a school level up to incomplete primary school (no more than 7 school years); level 2 refers to a school level up to graduation, if the occupation of the speaker does not need the university degree; level 3 refers to a higher school level. In the mini-corpus, conversations feature 4 speakers with school level 1, 11 with school level 2 and 8 with school level 3; dialogues feature 2 speakers with school level 1, 7 with school level 2 and 5 with school level 3; monologues feature 3 speakers with school level 1, 2 with school level 2 and 2 with school level 3.

The most important feature of the Brazilian mini-corpus is its large diaphasic variation. As one can see in Table 1, the mini-corpus includes many different communicative situations. The diaphasic variation is an important parameter, on one hand because it is what ensures that the texts are really spontaneous and produced in natural contexts, and on the other hand, because diaphasic variation leads to variation in the information structure and in illocutionary values within the corpus.

The Brazilian mini-corpus maintains the same structure of the informal C-ORAL-BRASIL, divided into two sections, family/private and public situations, which are subdivided into conversations, dialogues and monologues. As in informal language perfect monologues are almost impossible, monologues are here defined as situations in which there is a clear predominance of textual elaboration by one of the speakers and almost no interaction. Dialogues are situations in which the linguistic exchange is focused on two informants (even if there are more minor intervenients) that produce a text highly entrenched in the extra-linguistic context. Conversations are much like dialogues, but they involve the active participation of three or more speakers. Table 2 shows the word distribution in each branch of the Brazilian mini-corpus.

Table 2. Number and proportion of words of the Brazilian mini-corpus

Context	Total		Conversations		Dialogues		Monologues	
Total	31318	100%	9774	31%	11331	36%	10213	33%
Family/private	23272	74%	6348	20%	8325	27%	8599	27%
Public	8046	26%	3426	11%	3006	10%	1614	5%

The Brazilian mini-corpus has a total of 31,318 words in 3:58:36 of recording. The distribution of words in each branch of the mini-corpus is showed in Table 2. In total, there is a balance regarding the percentage of words in each type of interaction: conversations have 31% of words, dialogues have 36% and monologues have 33% of total words.

It is important to say that, for many aspects, conversations and dialogues should be considered as one interactive typology versus monologues, that are a textual typology; therefore, a balanced mini-corpus should endure 2/3 of interactional typology and 1/3 of textual typology. The family/private context comprises 74% of the total number of words, and texts in public contexts represent only 26% of the total words in the mini-corpus. Due to the low representativeness of the public context, it is not possible to consider the context as a variable in studies based in the Brazilian mini-corpus.

3. Informational tagging

All 20 texts received informational tagging, using the set of informational units proposed by the Language into Act Theory and the Informational Patterning Hypothesis (Cresti 2000). In this framework, each utterance can be analyzed informationally. The only unit that is necessary and sufficient to build an utterance is the Comment unit, since it carries the illocutionary force of the speech act and gives prosodic and pragmatic autonomy to the utterance. The complex utterances consist of the comment unit and one or more units that accomplish different functions. These unit can be textual, when their function is to build the very text of the utterance, or dialogic, when their function is to support the interaction. The textual units, besides the Comment, are Topic (TOP), Appendix of Comment (APC), Appendix of Topic (APT), Parenthetical (PAR) and Locutive Introducer (INT). The dialogic units are Incipit (INP), Conative (CNT), Allocutive (ALL), Phatic (PHA), Expressive (EXP) and Discourse Connector (DCT).

Each unit is identifiable through three criteria: a functional criterion, a prosodic criterion and a distributional criterion; so, each unit has its specific function, its specific prosodic profile and its specific or preferential position in the utterance.

Other complex informational patterns are formed by Multiple Comments (CMM). In these cases two or more comments in the same utterance produce a rhetorical effect that causes that the two (or more) speech acts are interpreted as a whole. This is what happens in lists, comparisons, reinforcement, confirmation requests, among others types of multiple comments (Raso in press). Sometimes an informational unit can be segmented into more than one tone unit, characterizing the phenomenon of the Scanning unit (SCA). The scanning unit is due to difficulty in speech production, emphatic reasons or to articulatory necessity in case of too extended information units in terms of syllabic dimensions.

Finally, when there is less actional and interactional activity and the speakers builds a semantic text, the utterance is somehow dilated, giving rise to what is called *Stanza*. *Stanzas* are linguistic entities that do not correspond to the execution of one illocutionary force nor of a conventionalized rhetoric pattern, but to a broader linguistic activity, such as the construction of narratives and arguments. The *stanzas* are composed of sequences of Bound Comments (COB), whose junction is processual and not patterned. A complete listing of informational tags are shown in Figure 1. Later on this paper we will deep in the description of each information unit.

Textual information units		**Dialogic information units**	
COM	Comment	INP	Incipit
CMM	Multiple Comment	CNT	Conativ
COB	Bound Comment	PHA	Phatic
COB_s	Subordinator Comment	ALL	Allocutive
TOP	Topic	EXP	Expressive
TPL(n)	List of Topic: n indicates ordinal sequence	DCT	Discourse Connector
TOP_s	Subordinator Topic		
APC	Appendix of Comment	**Informationally empy units**	
APT	Appendix of Topic	SCA	Scanning
PAR	Parenthetic	EMP	Empty (incomplete units)
PRL	List of Parenthetic	TMT	Time Taking
INT	Locutive Introducer	UNC	Non identifiable
Further mark			
_r	Reported speech unit		

Figure 1. Tagset for the information units

Before they start the tagging, the annotators went through a phase of training, exercises and discussions that involved the project coordinator and the researchers of the LABLITA lab. The goal was not only to enable annotators with respect to the theoretical tools, but also to establish a standard of uniformity and consistency. The annotators also went through a statistical evaluation of the degree of agreement before beginning the informational tagging task. All annotators independently tagged a dialogue with 120 utterances (171 tone/information units) and a monologue with 70 utterances (372 tone/information units).

The overall results of the inter-rater agreement test (Kappa Statistics) were 0.62 for the utterance and 0.73 for tone/information unit. A more detailed analysis pointed out that the disagreement cases were restricted to just a few information tags. These problems could be managed in the revision phase. Most of the problems encountered in the tagging and showed by the Kappa Statistics test involved the two specific information units: Multiple Comments and Bound Comments, which are the less studied units so far.

Tagging went through two distinct phases of review. The first was conducted by one annotator, always different from who had originally tagged the text, together with the coordinator of the project. Sometime later, the informational tagging was again reviewed by the project coordinator in conjunction with a member of the European project (C-ORAL-ROM), which is the most experienced person in relation to informational tagging based on the Language into Act Theory[4]. This last revision had both the goal to better the accuracy of the informational tagging and to ensure consistency with the tagging of the Italian corpus.

4. Structural and informational features

The first measurements to be observed in order to obtain a better knowledge of spontaneous speech are the distribution of dialogic turns, the number of utterances and the number of tone/information units in the sample and its branches. The averages of utterances per turns and of tone/information unit per utterance allow to evaluate the degree of interaction of the texts. The lower these numbers, the higher the interaction degree.

The average of utterances per turn is a measurement that reflects the alternation of the turns during the interaction: therefore, if the turns are short in terms of utterances, this means that the interactivity is high; when the turns show many and longer utterances, this reflects a lower degree of interactivity. As far as the average of tone/information unit per utterance is concerned, we can observe that the higher the number of tone unit per utterance is, the more complex the utterances are; a high number of very complex utterances is typical of interactions with a low degree of interactivity. The reason is that the utterance complexity goes together with the amount of textual information units; and the more text we have in the interaction, the less percentage of illocution, i.e. actionality, and therefore interactivity, we have. Conversations and dialogues show turns with a lower number of utterances and utterances with a lower number of tone/information units.

Table 3 shows these values for each text and for each interactional typology. The Table also presents other values: the average values of utterances per turns

[4] We thank Ida Tucci of the LABLITA lab for her collaboration.

calculated considering only the concluded utterances, and the average value of tone units per utterances calculated considering only the concluded tone units.

Looking at the data of Table 3, we can observe several characteristics of the texts and their structure. First of all, it is evident the difference between dialogues and conversation on one hand, and monologues on the other hand, in terms of number of turns average. While conversation and dialogues have almost the same number of turns (respectively 1333 and 1371), monologues show a much lower number of turn (250).

The same opposition between dialogues and conversations on one hand and monologues on the other hand can be confirmed with respect to other measurements, as already observed for the C-ORAL-ROM languages by Cresti (2005):

- The average of concluded utterances per turn is similar between conversations and dialogues (respectively 1,39 and 1,66), while it is much higher for monologues (3,68);
- The average of tone units per utterance also is similar for conversations and dialogues (1,71 and 1,54) and much higher for monologues (2,94);
- The number of retracting phaenomena is also similar for conversations and dialogues (253 and 228) and much higher for monologues (388).

Table 3. Structural features of Brazilian mini-corpus

Text typology	Dialogic turns (DT)	Interrupted sequences	Concluded sequences (CS)	CS/DT	Retracted units	Informative tone units	IU/CS
Total	2954	441	5043	1,71	869	9384	1,86
Conversations	1333	191	1848	1,39	253	3164	1,71
bfamcv01	159	41	207	1,3	46	441	2,13
bfamcv02	239	29	356	1,49	36	579	1,63
bfamcv03	185	10	296	1,6	38	467	1,58
bfamcv04	323	43	422	1,31	28	645	1,53
bpubcv01	265	32	323	1,22	35	611	1,89
bpubcv02	162	36	244	1,51	70	421	1,73
Dialogues	1371	176	2275	1,66	228	3513	1,54
bfamdl01	338	24	542	1,6	19	781	1,44
bfamdl02	176	35	247	1,4	56	453	1,83
bfamdl03	172	38	300	1,74	41	505	1,68
bfamdl04	123	9	244	1,98	11	367	1,5
bfamdl05	239	40	391	1,64	46	566	1,45
bpubdl01	158	14	262	1,66	32	407	1,55
bpubdl02	165	16	289	1,75	23	434	1,5
Monologues	250	74	920	3,68	388	2707	2,94
bfammn01	19	8	98	5,16	70	245	2,5
bfammn02	95	13	171	1,8	57	477	2,79
bfammn03	48	9	135	2,81	59	353	2,61
bfammn04	26	8	181	6,96	21	446	2,46
bfammn05	31	18	135	4,35	56	401	2,97
bfammn06	6	4	72	12	47	328	4,56
bpubmn01	25	14	128	5,12	78	457	3,57

These measurement allow us to establish a first opposition between dialogic texts (conversations + dialogues) and monologic texts. This opposition will be confirmed analyzing the information structure of these two major typologies. Nevertheless this does not eliminate completely the differences between conversations and dialogues.

First of all it is necessary to note that in number of words the two typologies are not perfectly balanced (since the most important balancing is due to the opposition between dialogic and monologic typologies): in the mini-corpus, we have 9843 word for conversations and 11371 words for dialogues (since we have 7 dialogues and 6 conversations). This does not reflect any significant difference in term of turn dimensions, as conversations have 7,38 words per turn and dialogues have 7,18 words per turn. But if we observe the number of interrupted sequences, we note that its rate (number of interrupted sequences divided for the number of words) is 1,94 in conversations and only 1,54 in dialogues. Similarly, the rate of retractings is 2,57 in conversations and only 2,0 in dialogues. This means that the higher competition for the turn in conversation causes a higher number of fragmentation phaenomena.

Another interesting difference is the higher rate of tone units per turn in conversations (1,71) with respect to dialogues (1,54). This difference seems to reflect the fact that in conversations it is easier to find parts in which one speaker articulates more complex utterances, but we have also to consider that in the mini-corpus we have 2 conversations without a specific activity performed by the speaker, which can also contribute to a less actional interaction.

In fact, different text typologies, specially the opposition between dialogic typologies and monologic typology, have important consequences on the information structure of spoken discourse. Table 4 shows some important values in order to distinguish the structure of conversations, dialogues and monologues.

The data presented in this Table was extracted through the search interface of DB-IPIC, a database in XML format implemented by Panunzi e Gregori (2011; also in this volume). It allows the study of information units in spoken corpora annotated according to the Information Patterning Theory (Cresti 2000; Moneglia & Cresti 2006; Scarano 2009).

We can observe that the highest level of difference is the percentage of presence of the three units of reference: utterance, illocutionary pattern and stanza. As the data show, clearly more than 80% of conversation and dialogue structure is built by utterances, 10% by illocutionary patterns and only a very little part by stanzas, which, moreover, are usually very simple, in term of structure. The differences between conversations and dialogues are very little, but we will come back to this later.

The most important aspect now is to note how different is the composition of the monologic typology. It features only 66% of utterances, 8% of illocutionary patterns but 25% of stanzas, which are often very complex. So we can say that

stanza is a reference unit typical of monologic texts, and this is a very important feature of the informational complexity of this typology.

Table 4. Informational features of the Brazilian mini-corpus

Informational typologies	**Conversations**		**Dialogues**		**Monologues**	
Total linguistic entities	1855	100,0%	2304	100,0%	950	100,0%
Total utterances	1534	82,7%	1972	85,6%	633	66,6%
Simple utterances *COM*	1095	71,4%	1452	73,6%	351	55,5%
Simple scanning utterances *COM + SCA, TMT, EMP*	91	5,9%	121	6,1%	63	10,0%
Compound utterances with dialogic units *COM + ALL, CNT, DCT, EXP, INP, PHA*	196	12,8%	232	11,8%	63	10,0%
Compound utterances with textual units *COM + APC, INT, TOP, TPL, APT, PAR, PRL*	108	7,0%	125	6,3%	100	15,8%
Mixed compound utterances *COM + textual and dialogic units*	44	4,0%	42	2,9%	56	16,0%
Total illocutionary patterns	202	10,9%	225	9,8%	77	8,1%
Simple illocutionary patterns *2 or more CMM*	147	72,8%	148	65,8%	34	44,2%
Simple scanning illocutionary patterns *2 or more CMM + SCA, TMT, EMP*	13	6,4%	19	8,4%	10	13,0%
Compound illoc. patterns with dialogic units *2 or more CMM + ALL, CNT, DCT, EXP, INP, PHA*	24	11,9%	30	13,3%	8	10,4%
Compound illoc. patterns with textual units *CMM + APC, INT, TOP, TPL, APT, PAR, PRL*	14	6,9%	20	8,9%	21	27,3%
Mixed compound illocutionary patterns *2 or more CMM + textual and dialogic units*	4	2,0%	8	3,6%	4	5,2%
Stanzas *at least one COB + COM*	119	6,4%	107	4,6%	240	25,3%

But the complexity of the monologic typology is also testified by analizing the internal structure of the utterance. In conversations and dialogues, the most part of the utterances are simple utterances, while in monologues the proportion of compound utterances is much higher.

It is interesting also to observe the percentage of simple scanning utterances, that means utterances built by the comment and one or more informationally empty units, like scanning units or time taking or not concluded units. It is important also to observe that the scanning simple utterances have a much higher weight in monologues than in the dialogic typology. This depends on at least two factors: first, the higher fragmentation phaenomena in monologues, due to the processual construction of a more complex text, and second, to the fact that many of these cases, certainly more than in dialogic typology, are due to the interruption of a compound unit, and therefore are included inside simple utterances only because the interruption happens before the realization of a full informational unit.

The monologic informational complexity can be confirmed by another important aspect: the relevance of textual units in building compound utterances. In Table 4 there is a differentiation among compound utterances with dialogic units, compound utterances with textual units and mixed compound utterances. This last category includes all the compound utterances that have both textual and dialogic units inside. For our purpose here, utterances that have textual units, independently if they have also dialogic units or not, will be considered as one unified category and compared with the compound utterances with only dialogic unit (besides, of course, the comment unit). Compound utterances with only dialogic units are more frequent in conversations and dialogues, where they sum respectively 12,8% and 11,8% of all the utterances, that are respectively 82,7% and 85,6% of all the reference units. Only 11,% of the utterances in conversations and 9,2% in dialogues are compound utterances with at least one textual unit. In monologues, what happens is much different: just 10% of the compound utterances are build only by dialogic units, while 31,8% have at least one textual unit.

If we now analyze the illocutionary patterns, we realize that they are more frequent in dialogic typologies, but also that they are more complex in textual typologies. In fact, without considering the simple scanning illocutionary patterns (that may depend on different reasons), we can observe that only 9,% of the illocutionary patterns in conversation and 12,5% in dialogues have textual units, while in monologues illocutionary patterns with textual units reach 32%.

All these measurements allow us to conclude that dialogic typologies are basically built on a sequence of simple utterances or illocutionary patterns. This means that these typologies are strongly based on alternation of the illocutionary force. The high presence of dialogic units shows that if the speaker needs more units than the illocutionary ones, they are still directed to the interlocutor in order to guaranty the interaction (dialogic units), and do not build the text of the utterance. The presence of textual units is in fact very low. On the contrary all the measurements in Table 4 lead us to conclude that monologic typology has a completely different structure. The very high weight of stanzas and of compound utterances with textual units shows that the importance of the illocutionary force is

much lower, while the importance of really informative units, that means units that build the text of the reference unit, is very high.

This can be explained with the fact that the basic activity the speakers perform when interacting and when build a monologic text is different: the interaction is an alternation of actions that the speakers do toward their interlocutor, and for this they need, besides the illocutionary force, also dialogic unit that provide the regulation of the channel and that of the social cohesion between the speakers. On the other hand, monologues are principally elaborated texts (argumentations, narratives, explanations, descriptions) built by only one speaker. He may have a certain degree of interaction with the listener(s), but his activity is mainly that of organizing and giving voice to his thinking, not to perform actions pulsioned during the interaction.

While the dialogic texts develop on the basis of interaction, monologues are a process of text construction by just one speaker. In dialogic texts the speaker does not have a mental project to develop, and interacts with the interlocutor depending on unforseeable interlocutor's action. In monologues the speaker does have a mental project, for instance to tell a story or to explain something, and this lead to a complex mental process in which the illocutionary force weakens and the semantic text construction takes, to a certain extent, its place.

5. Information structure in Brazilian Portuguese and in Italian

5.1 The characteristics of the information units

Before making a very general comparison between the Brazilian and Italian mini-corpora, it is necessary to offer some more informations about the function, the prosodic profile and the distribution of the information units[5].

The textual units build the text of the utterance. The only unit that is necessary and sufficient to build an utterance, as it carries the illocutionary force, is the Comment unit (COM). When this unit is patternized with another illocutionary unit gives rise to the Multiple Comment (CMM). In prosodic terms, they are root units ('t Hart et al. 1990), and are the only unit that has prosodic and pragmatic autonomy. Its prosodic profile changes according to the illocution that is conveyed (Firenzuoli 2003; Moneglia 2011) and always bears a functional nucleus, that is the prosodic portion that conveys the specific illocutionary value (see Mello & Raso in this volume; Cresti in this volume). Its distribution is free. Also the Bound Comments COB) are root units, but with a weakened illocutionary value. They appears in

[5] More detailed informations about the different information units can be accessed in Cresti (2000), Raso (in press) and in the bibliography about the specific unit.

stanzas (Cresti 2009) and are typical of monologic text, where *stanzas* can be very big and complex, organized in subpattern around each Bound Comment. The Bound Comment ends with a continuity prosodic signal, that marks that the reference unit is not concluded and that the illocutionary force must be interpreted inside a broader reference unit.

Figure 2 shows the distribution of the different root units in the three text typologies in the Brazilian mini-corpus. Once again we notice that the different unit have a similar distribution in conversations and dialogues, but a very different one in monologues.

While the greatest part of root units for conversations and specially dialogues is the Comment unit, for monologues Bound Comments have a very important role, reaching almost 1/3 of all the root units.

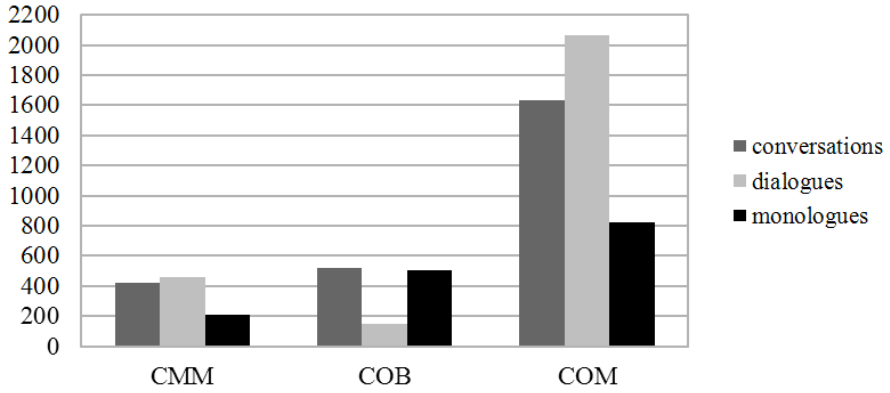


Figure 2. Distribution of the root units in the three text typologies

Actually, the real weight of Bound Comments in monologues is much bigger than Figure 2 shows. In fact, it is very common that the interlocutor constantly signals his attention by uttering simple utterances like hum hum // or exclamations that show his participations in the interaction. All these cases, which should not be considered within the monologue structure, are computed in the graphic as Comment unit. On the contrary, the weight of Bound Comments in conversations is only 20% and in dialogues 5,5%. As far as Multiple Comments are concerned, they concentrate in the dialogic typologies. Comparing conversations and dialogues, it is possible to observe that conversations have a little less Comment units and a higher presence of bound comments.

A compound information pattern contains one (or more) root units and normally has also textual or dialogical information units.

The textual information units are:

- The *Topic* (TOP) unit (Signorini 2003) is the most important unit of an information pattern. Its function is to define the cognitive perspective, that means the semantic dominion, of the illocutionary force. Prosodically, it is the only unit, besides the comment, that bears a functional nucleus, despite the fact that, like all information units except the comment, it is not pragmatically interpreTable in isolation. The nucleus is always, entirely or partially[6], positioned on the right of the unit (Firenzuoli & Signorini 2003; Raso et al. forthcoming). Its distribution is always on the left of the comment.
- The *Appendix* unit integrates the text of the Comment (Appendix of Comment – APC) or of the Topic (APT). Prosodically the Appendix has a descendent or flat profile. The APT can show movement, but without any focus. Their distribution is always on the right of the Comment or of the Topic (Raso & Ulisses 2008; Ulisses 2008; Tucci 2006).
- The *Parenthetic* (PAR) has the metalinguistic function to make a commentary about the utterance or part of it. Its profile is flat, with a lower (or rarely higher) F0 level with respect of the rest of the utterance, and a higher speech rate. It can occupy any position, even inside another textual unit, except the beginning of the utterance (Tucci 2004; Tucci 2009).
- The *Locutive Introducer* (INT) has the function to introduce a list of topics and specially an illocutionary pattern with a meta-illocutionary value, outside of the deictic coordinates of the utterance (Corsi 2009; Maia Rocha 2010; Maia Rocha & Raso 2011). One very important function of the INT is therefore that of marking the suspension of the pragmatic coordinates of the utterance introducing a different *hic et nunc*. Prosodically, INTs have a descendent profile, with a much lower F0 frequency with respect to the meta-illocution that follows, producing a clear F0 contrast that marks also prosodically the suspension of the pragmatic coordinates, and with a much higher speech rate. Its distribution is before the introduced units.

Figure 3 shows the distribution of the different textual units in the three text typologies of the Brazilian mini-corpus.

[6] Some Topic prosodic profiles have two semi-nuclea. In this case, the preparation (that depends on the syllabic dimensiono f the locutive contet) can be positioned between the two nuclear portions.

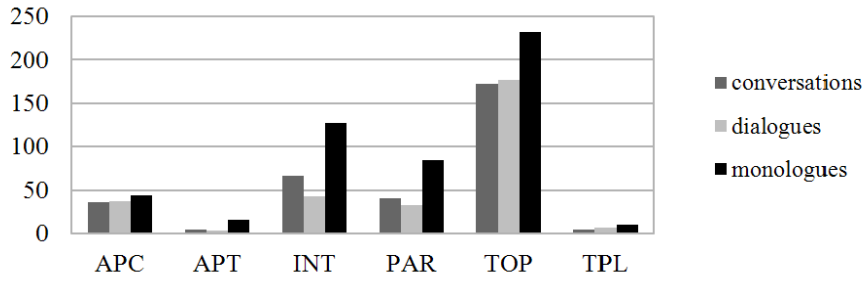


Figure 3. Distribution of the textual units in the three text typologies

It is noticeable that all textual units have a much higher presence in monologues. This is specially true for Topics, that are much more necessary in situations that cannot have the pragmatic situational context as an immediate reference for the illocutions, like in narratives, descriptions or argumentations, for the Locutive Introducer, since in monologues it is much higher the use of meta-illocutions, and for parenthetical, that allows the speaker to modalize and to make commentary on the textual content of the utterance.

Again, it is possible to notice a small difference between conversation and dialogues, always with conversations showing, in a very little proportion, the tendency to present some characteristics of monologues.

The dialogic units (Frosali 2008) are very different, with many respects, from the textual ones. Their function is not that to build the text of the utterance, but that of controlling the interaction. The dialogic units are:

- The *Incipit* (INP) has the function of beginning the turn or the utterance with contrast with the previous one; its prosodic profile is ascendent-descendent (or only ascendent or only descendent) reaching a high F0 value with a very short duration and high intensity; it opens the utterance.
- The *Phatic* (PHA) has the function to signal that the channel is open, with a very short and flat or descendent profile, and with low intensity; its position is free.
- The *Allocutive* (ALL) has two functions: to individualize the interlocutor, but specially to mark the social cohesion with him; its prosodic profile is descendent or slightly modulated, with standard duration and intensity; it must not be confused with the recall illocution (Raso & Leite 2010).
- The *Expressive* (EXP) has the function to support emotionally the illocution; its profile may vary, but it is usually modulated, with standard duration and intensity.

– The *Conative* (CNT) has the function to press the interlocutor to do or quit doing something; its profile is descendent, with short duration and high intensity.
– The *Discourse Connector* (DCT) has the function to open the utterance without contrast with the previous one, or to connect the subpatterns inside a *stanza*; its profile is flat or modulated, with high intensity and long duration.

Figure 4 shows the distribution of dialogic units in the three text typologies of the C-ORAL-BRASIL mini-corpus. The distribution of the dialogic unit is very interesting to show some specific aspects of the three text typologies. First, we can observe that is frequent an opposition between dialogic typologies and monologues. This is clear with respect to Conatives, Expressives and Discourse Connectors. In the first two cases, the dialogic typologies show a clearly higher presence of these units, but for discourse connectors the opposite happens. We will be back on this later.

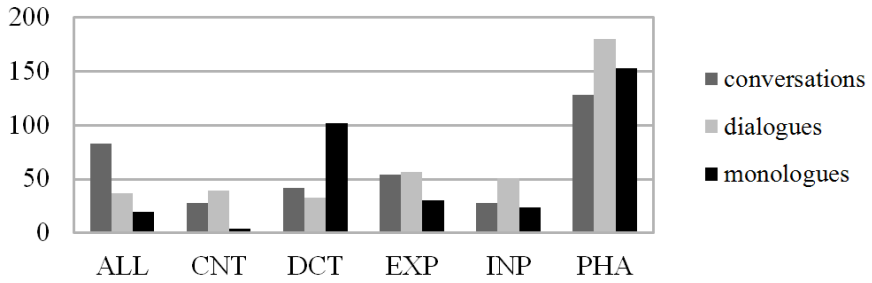


Figure 4. Distribution of the dialogic units in the three text typologies

Concerning the Allocutives, there is a sort of scale that goes from the highest presence in conversations to the lowest presence in monologues. This distribution of this unit depends on some well-studied factors. Allocutives are a strongly dialogic unit, as they are used always to support the interaction. They have, as already said, two main functions: that of individualizing the interlocutor and of marking the social cohesion with him. This last function is equally strong in conversations and dialogues, but very low in monologues. The first function does not make sense in dialogue but only in conversation. So this explain the fact that conversation has a higher use of allocutives with respect to dialogue. But what is the function of allocutives in monologues? Monologues, specially narratives, have a high amount of reported speech; in reported speech allocutives are used to indirectly signal to the interlocutor who are the reported speaker and the reported interlocutor.

The distribution of incipit and phatic still needs to be studied.

The distribution of Discourse Connector reflects the specific function of this unit, for many respects different from the other dialogic unit. As we said, its function is to mark continuity between two utterances, but also to connect subpatterns in a *stanza*. As the former function is common to the three text typologies, the last one is typical of the monologic typology, where *stanzas* are much more present and much more complex.

5.2 A first comparison with the Italian tagged mini-corpus

The informational tagging of the Italian C-ORAL-ROM corpus began much earlier than the tagging of C-ORAL-BRASIL. Therefore, in order to study the information structure in a cross-linguistic perspective, part of the Italian tagged corpus was extracted to be compared with the 20 tagged texts of the C-ORAL-BRASIL corpus. As the Brazilian mini-corpus is highly actional, to turn the Italian mini-corpus comparable to the Brazilian one, the priority was to maintain the same proportion between dalogic and monologic typologies and to maximize the actionality of the text, meaning with this, the maximum number of varieties of activities performed by the speaker while interacting. The composition of the Italian mini-corpus is that presented in Table 5.

The Italian mini-corpus is a little bigger, in terms of words, than the comparable Brazilian one, but its balancing, with respect to the two priorities (1/3 of monologic and 2/3 of dialogic texts, and maximization of different actional texts) is almost perfect. Since the Italian mini-corpus was adapted to the Brazilian mini-corpus, it cannot maintain the almost perfect balance with respect to the speakers' characteristics.

Here, we will only propose some general observations comparing the two mini-corpora. A better and deeper comparison needs a specific dedicated study. Table 6 shows for Italian the same data that Table 4 shows for Brazilian.

Table 5. The Italian mini-corpus

Text	Situation	M	F	Words
Total		23	31	34208
Conversations		9	11	10141
ifamcv01	relatives talk while browsing through family photos	1	2	
ifamcv09	friends explain the game Mastermind	3	0	
ifamcv15	family talks with child during lunch preparation	2	3	
ipubcv01	exchanging ideas during a meeting of a voluntary association	1	4	
ipubcv05	chat in a ironmonger while shopping	2	2	
Dialogues		5	13	12435
ifamdl04	interview of an artisan in his leather workshop	1	2	
ifamdl12	friends at home making a cake	0	2	
ifamdl15	beautician and customer in the beauty-center	0	2	
ifamdl17	two friends develop photos in a dark-room	1	1	
ifamdl19	father gives driving lesson to his daughter	1	2	
ifammn17*	professional explanation to a colleague about office-work	0	2	
ipubdl02	proposal of an insurance policy	0	2	
ipubdl05	teachers' meeting at the school office	2	0	
Monologues		9	7	11632
ifammn02	interview with an old partisan at his home	2	0	
ifammn05	elderly woman tells life story to her relatives	1	2	
ifammn08	narrative to a relative about the honeymoon	0	1	
ifammn03	an after-dinner travel tale to friends	2	2	
ifammn14	interview with a retired travelling-salesman	1	1	
ipubmn01	political speech at a political-party meeting	2	0	
ipubmn04	interview with an employee of the Poggibonsi municipality	1	1	

*Labeled as monologue but is acctually a dialogue

Table 6. Information features of the Italian mini-corpus

Informational typologies	**Conversations**		**Dialogues**		**Monologues**	
Total linguistic entities	1769	100,0%	2054	100,0%	1195	100,0%
Total utterances	1481	83,7%	1714	83,4%	842	70,5%
Simple utterances *COM*	987	66,6%	1169	68,2%	329	39,1%
Simple scanning utterances *COM + SCA, TMT, EMP*	95	6,4%	126	7,4%	90	10,7%
Compound utterances with dialogic units *COM + ALL, CNT, DCT, EXP, INP, PHA*	144	9,7%	178	10,4%	116	13,8%
Compound utterances with textual units *COM + APC, INT, TOP, TPL, APT, PAR, PRL*	172	11,6%	168	9,8%	186	22,1%
Mixed compound utterances *COM + textual and dialogic units*	83	8,4%	73	6,2%	121	36,8%
Total illocutionary patterns	183	10,3%	172	8,4%	80	6,7%
Simple illocutionary patterns *2 or more CMM*	106	57,9%	93	54,1%	25	31,3%
Simple scanning illocutionary patterns *2 or more CMM + SCA, TMT, EMP*	23	12,6%	17	9,9%	10	12,5%
Compound illoc. patterns with dialogic units *2 or more CMM + ALL, CNT, DCT, EXP, INP, PHA*	15	8,2%	22	12,8%	14	17,5%
Compound illoc. patterns with textual units *CMM + APC, INT, TOP, TPL, APT, PAR, PRL*	31	16,9%	28	16,3%	21	26,3%
Mixed compound illocutionary patterns *2 or more CMM + textual and dialogic units*	8	4,4%	12	7,0%	10	12,5%
Stanzas *at least one COB + COM*	105	5,9%	168	8,2%	273	22,8%

We can confirm that for Italian, dialogic texts behave in similar way, while monologic texts present very different measures. We can observe that the proportion of utterance is the same comparing dialogues and conversations. This allows us to hypothesize that the small differences found between these two typologies in the Brazilian mini-corpus are due to the presence of two conversations in which the speakers do not perform any specific activity, pushing therefore some measurements

in the direction of monologic values. We can also observe that monologues in the Italian mini-corpus present a little less stanzas and more utterances, but also a little less illocutionary patterns. In any case, these can be considered not to be significant differences.

A more significant difference is the fact that in Italian the percentage of simple utterances is much lower than in Brazilian. While Brazilian shows 71.4% of simple utterance in conversation, 73.6% in dialogue and 55.5% monologue, in Italian these measurements are, respectively, 66.6%, 68.2% and 39.1%, what seems to lead to a more complex informational structure for this language. This hypothesis is strengthened by the fact that the number of textual compound utterances is also higher in Italian. While Brazilian shows a percentage of 11.00%, 9.2% and 31.8% of textual compound utterances respectively for conversations, dialogues and monologues, Italian presents 20.0%, 16.00% and 58.9%. The same happens for illocutionary patterns: compound illocutionary patterns are much more common in Italian, while simple illocutionary patterns are much more common in Brazilian. Differences in terms of stanzas do not seem significant. Figure 5 shows the proportion of root units in the Italian mini-corpus.

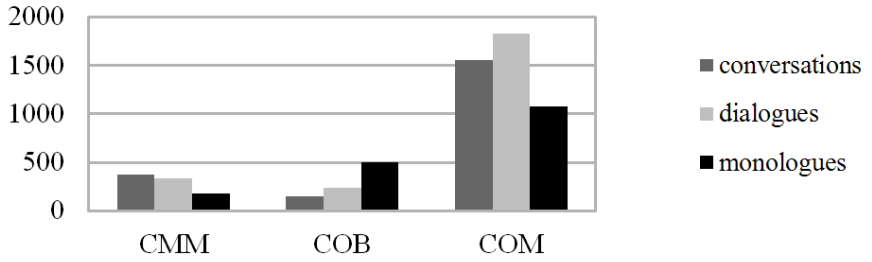


Figure 5. Distribution of the root units in the three text typologies in Italian

With respect to Brazilian root units and its distribution in the different branches of the mini-corpus, it is noticeable a lower number of illocutionary patterns: 10.3%, 8.4% and 6.7% respectively in conversations, dialogues and monologues, versus 10.9%, 9.8% and 8.1% in Brazilian. On the contrary, the number of bound comments is much higher (with the exception of conversations).

Figure 6 shows the distribution of textual units in the Italian mini-corpus and corresponds to Figure 3 for Brazilian. We can observe the much higher number of all textual units in Italian, with the only exception of the Locutive Introducers.

The fact that Locutive Introducers are in contratendential distribution with respect of the other textual units is something that must be explained: first of all we can observe that in Italian the INTs distribution does not vary much in the three typologies, even if monologues have more INTs and dialogues have less INTs; in

the Brazilian mini-corpus the number of INTs in monologues is much higher than in the other typologies.

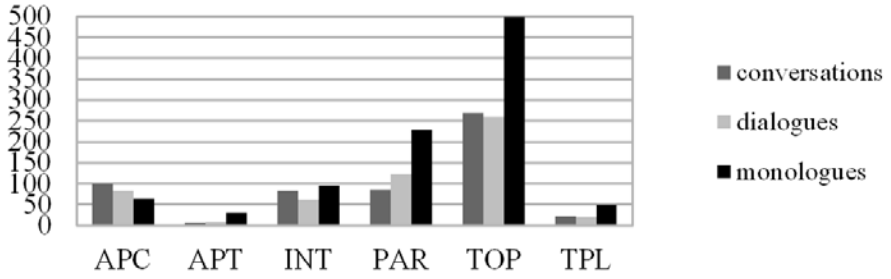


Figure 6. Distribution of the textual units in the three text typologies in Italian

A hypothesis that should be tested is that reported meta-illocutions, and specially reported speech, are much more frequent in Brazilian, since they represent a more pragmatic and less textual strategy of text building. Another interesting difference between the two mini-corpora with respect to textual units is the inverted distribution in the different typologies of the APCs. While Brazilian has more APC in monologues and less in conversations, Italian shows more APCs in conversations and less in monologues.

Figure 7 shows the distribution of the dialogic units in the Italian mini-corpus, and corresponds to Figure 4 for Brazilian.

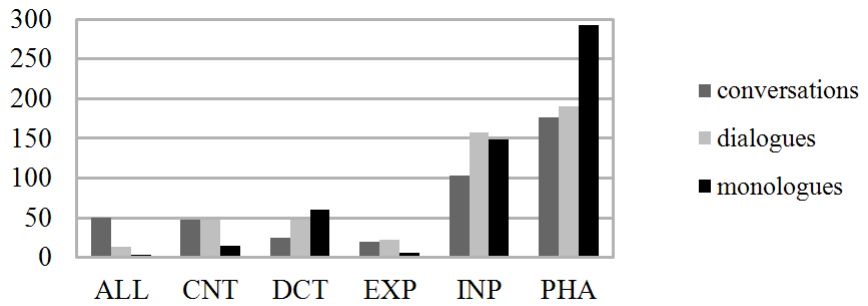


Figure 7. Distribution of dialogic units in the three text typologies in Italian

The different distribution of dialogic units in the two mini-corpora allows for many considerations. First of all it is important to emphasize the cultural relevance of dialogic units. They have the function to govern the interaction, and this is a very sensible to cultural characteristics function.

A study about allocutive in Italian, Spanish, European Portuguese and Brazilian Portugues (Raso & Leite 2010) shows that Brazilian Portuguese and European

Portuguese have a very different way to use this unit, with a difference between them higher than the difference that they show with respect to Spanish and Italian. Comparing Brazilian and Italian with respect to all the dialogic units, we note that Brazilian uses much more allocutives and expressives, while Italian uses much more conatives and incipits. The very high presence of phatics in Italian monologues is another remarkable difference. The last difference is the very high number of DCTs in monologues. These differences should still be better studied.

6. Annotation of prosodic boundaries and informational tagging

In this section we discuss the relationship between the annotation of prosodic boundaries and the identification of the informational value for the prosodic units in the Brazilian mini-corpus[7]. This research is necessary to the extent that the Brazilian mini-corpus transcripts had not undergone a revision after the text-to-speech alignment as the rest of the C-ORAL-BRASIL. Thus, during the informational tagging, annotators add or remove either words or prosodic breaks. In several cases, they also change the value of a prosodic break (for instance, from terminal to non-terminal or *vice versa*). Thus, this analysis aims to assess to what extent these changes were made in the prosodic annotation during tagging, and also discuss the change in the annotation with relation to specific information functions.

For this analysis we used two versions of the Brazilian mini-corpus. The first version consists of the transcripts after they passed through a first revision. The second one is the final informationally tagged version of the Brazilian mini-corpus. The total of analyzed transcripts amounts 40 texts. Each text went through an automatic processing through R computational tool (R Development Core Team 2010) in order to be prepared for data mining and statistical analysis. In a spreadsheet, each first version transcript was aligned, word by word, with the corresponding second version transcript. Naturally the versions of each text presented a different word numbers, due to word inclusions and exclusions in the final version. As the inclusion or exclusion of words can alter the annotation of prosodic breaks, changes in transcripts at the segmental level were also controlled.

After alignment, the sample adds up to a total of 31,750 tokens. Each token corresponds to a word boundary, considering the words of both versions. Of this total, 11,200 (35%) positions had a prosodic break either in the first or in the second version. Considering only these positions, we noticed that 6% of the tokens (651 cases) are involved in some sort of alteration in the segmental level, like additions, deletions and corrections of words (see Table 7).

[7] For a detailed description of the methodology for segmentation e its validation in the C-ORAL-BRASIL corpus, see Raso & Mittmann (2009), Mello et al. (in press).

Table 7. Types and frequencies of positions with annotation of prosodic breaks

Position type	**Freq.**	**%**
Total positions with prosodic breaks	11200	100%
Positions with segmental changes	651	6%
Word corrections	365	3%
Word inclusions	213	2%
Word exclusions	73	1%
Positions without segmental changes	10549	94%
Without changes in prosodic breaks	9175	82%
With changes in prosodic breaks	1374	12%

We do not consider for the analysis the positions in which there was any kind of modification at the segmental level. In this way, we eliminate possible changes in the annotation of prosodic breaks due to additions or deletions of words. Thus, the total analyzed data equals 1,374 tokens. Those correspond to the instances in which, at the same time, there were no segmental changes but that presented changes on the annotation of prosodic breaks.

As shown in Table 7, during the informational labeling, annotators made changes in 12% of the prosodic breaks. This value is high, nevertheless we must take into account that the transcripts underwent only one phase of revision before informational tagging, while the rest of the C-ORAL-BRASIL informal corpus passed by at least 4 revisions.

Changes include the addition and deletion of prosodic breaks, as well as the modification of the prosodic breaks value. The changes made during the informational tagging are summarized in Table 8.

Considering break exclusions (26% of total changes), one can notice that an irrelevant percentage of those relates to terminal breaks (0.29%) and to retracting and interruption (both equals 0.95%). Almost all the exclusions consist of non-terminal breaks deletions (24.09% and 331 cases).

Table 8. Types and frequencies of prosodic break changes

Type of prosodic break change	Freq.	%
Total positions with changes in prosodic breaks	**1374**	**100.00%**
Exclusion of prosodic break	**361**	**26.27%**
Terminal	4	0.29%
Non-terminal	331	24.09%
Interruption	13	0.95%
Retracting	13	0.95%
Inclusion of prosodic break	**375**	**27.29%**
Terminal	11	0.80%
Non-terminal	355	25.84%
Interruption	4	0.29%
Retracting	5	0.36%
Modification of prosodic break type	**638**	**46.43%**
Terminal → non-terminal	354	25.76%
Terminal → interruption	32	2.33%
Terminal → retracting	2	0.15%
Non terminal → terminal	90	6.55%
Non terminal → interruption	19	1.38%
Non terminal → retracting	24	1.75%
Interruption → terminal	27	1.97%
Interruption → non terminal	14	1.02%
Interruption → retracting	45	3.28%
Retracting → terminal	5	0.00%
Retracting → non terminal	22	1.60%
Retracting → interruption	4	0.29%

Most non-terminal break deletions (around 57%) are due to the inappropriate association of prosodic boundaries and discourse markers. Examples (1) and (2) below illustrate such occurrences.

(1) então / vamo passar lá // (bfamdl05) first version
então vamo passar lá // (bfamdl05) final version
[so let's go there]

(2) mas é isso aí / o' // (bfammn01) first version
mas é isso aí o' // (bfammn01) final version
[so this is it see]

What happens is that many lexical items, especially in initial position in the utterance, are candidates to be discourse markers, like 'então' (*so*), 'aí' (*so*), 'mas'

(*but*), 'e' (*and*) and several other items. These are all items with low phonetic consintency, that may be realized very quickly; after them it is possible, but not necessary, that a prosodic break is realized, giving to them the status of discourse markers. As they are not syntactically compositional with the rest of the utterance, it is very likely that a boundary is perceived and attributed to prosodic aspects even when there is not any prosodic reason for this. This represents the typical case in which revisions reduce a wrong annotation.

Other significant exclusions (14%) are related to the false association between prosodic units and syntactic units. Non-terminal breaks were removed from the final version in contexts where a syntactic limit, such as clause ending, was falsely interpreted as containing also a prosodic boundary. See examples (3) and (4) below.

(3) eu ditando / e o Tommaso escrevendo // (bfamdl01) first version
eu ditando e o Tommaso escrevendo // (bfamdl01) final version
[I dictating and Tommaso writing]

(4) essa é a rua / que nós vimo // (bfamdl05) first version
essa é a rua / que nós vimo // (bfamdl05) final version
[this is the street that we saw]

The results show that almost all changes were related to non-terminal prosodic breaks. This is important for two reasons:

- non terminal breaks are less relevant in terms of perception; therefore, the fact that almost all problems in segmentation, after only one revision, were related to them means that the original segmentation and the first revision had been accurate;
- non-terminal breaks are precisely the prosodic breaks that relate to the realization of complex informational patterns in utterances, as well as the formation of stanzas and illocutionary patterns.

The proportion of changes according to each type can be better observed in Figure 8. Black slices indicate changes that originate terminal prosodic breaks, gray slices indicate the proportion of changes that create non-terminal breaks, and the hatched portions indicate changes that originate prosodic breaks with no informational value, i.e., retractings and interruptions.

It is clear that the insertion of non-terminal breaks and the switching of terminal breaks to non-terminal breaks are the major changes that must be understood. That is possible if we cross-tabulate the data of these two variables with the information tag that was assigned to them.

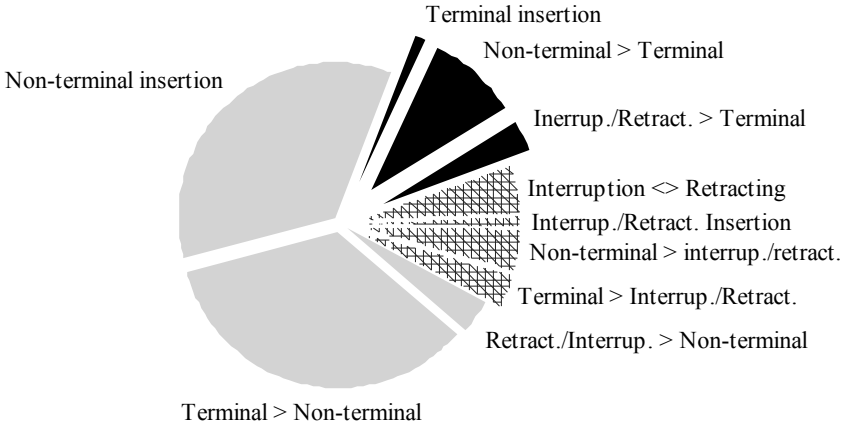


Figure 8. Proportion of different types of changes in prosodic annotation during informational tagging

Table 9 shows the total number of occurrences for each informational tag used in the informationally tagged Brazilian mini-corpus, the total number of changes in prosodic annotation associated with each tag and, also for each tag, the more detailed number of non-terminal breaks insertions and terminal to non-terminal breaks switchings.

These data allow us to see that most switches from terminal to non-terminal break concern the identification of Multiple Comments (CMM) forming illocutionary patterns and Bound Comments (COB) that form stanzas. It is, in fact, difficult sometimes to interpret the value of the prosodic break in cases like these, particularly during the transcription phase, but also during the revision of transcripts that are not aligned with the corresponding audio.

The terminal to non-terminal switching related to COB units reveals that the text-to-speech alignment improves the ability to make refined distinctions about prosodic break values. The annotator can more easily distinguish sequences of units with weak illocutionary value (stanzas) from those that really bear a conclusive prosodic value.

Also the recognition of many illocutionary patterns are facilitated by text-to-speech alignment. In many cases, each root unit (CMM) that composes the illocutionary pattern seems to function in isolation. During the informational tagging, text-to-speech alignment allows the annotator to have the perception of the rhetorical effect created by the units when considered together as part of a unique compound illocutionary pattern. Probably, most of the cases of recognition of illocutionary patterns need the cognitive perspective provided by informational tagging.

Table 9. Cross-tabulation between information tag and change in prosodic breaks annotation

Information tag	Total tokens	Tokens with prosodic annotation changes		Non-terminal break insertion		Terminal to non-terminal switching	
COM	4514	166	3.68%	24	14.46%	28	16.87%
CMM	1095	161	14.70%	55	34.16%	**91**	**56.52%**
COB	836	204	24.40%	57	27.94%	**136**	**66.67%**
TOP	581	132	22.72%	**106**	**80.30%**	9	6.82%
EMP	877	86	9.81%	0	0.00%	0	0.00%
SCA	914	79	8.64%	44	55.70%	0	0.00%
PHA	461	47	10.20%	9	19.15%	32	68.09%
PAR	152	30	19.74%	9	30.00%	16	53.33%
INT	236	24	10.17%	7	29.17%	12	50.00%
DCT	177	21	11.86%	17	80.95%	0	0.00%
INP	103	15	14.56%	9	60.00%	5	33.33%
EXP	141	9	6.38%	5	55.56%	4	44.44%
CNT	71	9	12.68%	1	11.11%	8	88.89%
TMT	139	6	4.32%	1	16.67%	1	16.67%
ALL	140	5	3.57%	0	0.00%	4	80.00%
APC	117	5	4.27%	0	0.00%	3	60.00%
APT	23	4	17.39%	4	100.00%	0	0.00%
UNC	53	2	3.77%	0	0.00%	0	0.00%
TPL	22	2	9.09%	2	100.00%	0	0.00%
i-COB	13	2	15.38%	2	100.00%	0	0.00%
i-COM	20	1	5.00%	1	100.00%	0	0.00%
PRL	6	1	16.67%	0	0.00%	1	100.00%
i-CMM	2	1	50.00%	1	100.00%	0	0.00%
i-TPL	1	1	100.00%	1	100.00%	0	0.00%
i-TOP	2	0	0.00%	0	0.00%	0	0.00%
i-PAR	1	0	0.00%	0	0.00%	0	0.00%
Total	10697	1013	9.47%	355	35.04%	350	34.55%

On the other hand, most of non-terminal insertions are linked to the identification of Topic units (TOP). Although this cases are more unexpected and difficult to explain, since Topics are signaled, in principle, with prosodic boundaries of high perceptual salience, two hypotheses can be raised to try to understand why transcribers did not perceive so many prosodic boundaries.

The first one has to do with the fact that a new prosodic profile of Topic was identified during the informational tagging. It is possible that the transcriber's perception was, to some extent, biased by the types of prosodic movements they expected to find. Thus, an unforeseen prosodic movement may have caused the transcribers to disregard it as a prosodic boundary signal. The second hypothesis is that transcribers may have missed non-terminal breaks associated with the border of Topic units when Topics coincide with the subject of the sentence. It is usual that the subject is produced with some prosodic prominence that signals its semantic prominence. Topics, differently, have a prosodic focus that signals its pragmatic

prominence, that is to instantiate a cognitive reference for the interpretation of the speech act. It is possible that less experienced transcribers may interpret a Topic as a subject and then miss to annotate the prosodic boundary. Anyway, this is a case that needs further research.

7. Final remarks

This paper presented for the first time two comparable mini-corpora for cross-linguistic analysis of information structure. The two compared languages are Brazilian Portuguese and Italian.

Giving only an overall look to the informational characteristics, it was possible to note some aspects that seem to be language independent, like the basic structure of the three different textual typologies, and some characteristics vary according to the language. A very important difference seems to be the tendency of Brazilian Portuguese to use much less textual units and to be more actional and less textual than Italian. At the same time, we observed that one textual unit, the locutive introducer, is much more used in Brazilian; we proposed an hypothis that could account for this particular feature and that would confirm the general characteristics observed for the different language strategies.

Another important aspect that the two comparable mini-corpora allows us to observe is the completely different behavior of the two languages with respect to dialogic units. These units are a very important feature to study sociolinguistic differences in cross-linguistic verbal behavior.

The last part of the paper aims to show how a different perspective (cognitive versus perceptual) can change the segmentation of the speech flow. The finding of this part of the study can have methodological consequences in speech segmentation, and can help to understand what is more or less salient for perception.

References

Corsi, G. 2009. L'introduttore locutivo: una ricerca corpus-based di italiano parlato informale. BA thesis, Università degli Studi di Firenze.

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

Cresti, E. & Moneglia, M. 2010. Informational patterning theory and the corpus-based description of spoken language: The compositionality issue in the topic-comment pattern. In M. Moneglia & A. Panunzi (eds), *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*. Firenze: Firenze University Press.

Cresti, E. 1987. L'articolazione dell´informazione nel parlato. In *Gli italiani parlati: sondaggi sopra la lingua di oggi*. Firenze: Accademia della Crusca, 27-90.

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

Cresti, E. 2005. Notes on lexical strategy, structural strategies and surface clause indexes in the C-ORAL-ROM spoken corpora. In E. Cresti & M. Moneglia (eds), *C-ORAL-ROM: Integrated reference copora for spoken Romance Languages*. Amsterdam/Philadelphia: John Benjamins, 209-256.

Cresti, E. 2006. LABLITA Corpus of Spontaneous Spoken Italian. http://lablita.dit.unifi.it/corpora/descriptions/lablita/

Cresti, E. 2009. Unità di analisi testuale e caratteri costruttivi nell'italiano parlato (spontaneo) e scritto (letterario). Ricerche corpus-based. In A. Ferrari (ed.) *Sintassi storica e sincronica dell'italiano. Subordinazione, coordinazione, giustapposizione*. Firenze, Cesati, vol. 2, 713-732.

Cresti, E., Panunzi, A. & Scarano, A. 2005. The Italian corpus. In E. Cresti & M. Moneglia (eds), *C-ORAL-ROM: integrated reference corpora for Spoken Romance Languages*. Amsterdam/Philadelphia: John Benjamins, 71-110.

Firenzuoli, V. 2003. Le Forme Intonative di Valore Illocutivo dell'Italiano Parlato: Analisi Sperimentale di un Corpus di Parlato Spontaneo (LABLITA). PhD diss., Università di Firenze.

Firenzuoli, V. & Signorini, S. 2003. L'unità informativa di topic: correlati intonativi. In G. Marotta (ed.), *La coarticolazione: Atti delle XIII giornate di studio del Gruppo di Fonetica Sperimentale*. Pisa: ETS, 177-184.

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

Frosali, F. 2008. Le 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 University Press, 417-424.

't Hart, J., Collier R. & Cohen, A. 1990. *A Perceptual Study on Intonation*: *An Experimental Approach to Speech Melody*. Cambridge: Cambridge University Press.

MacWhinney, B. 2000. *The CHILDES Project*: *Tools for Analyzing Talk.* 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Maia Rocha, B. 2010. A Unidade Informacional de Introdutor Locutivo no Português Brasileiro: uma análise baseada em corpus. BA thesis, Faculdade de Letras, Universidade Federal de Minas Gerais.
http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/DAJR-8ELJXZ

Maia Rocha, B. & Raso, T. 2011. A unidade informacional de Introdutor Locutivo no português do Brasil: uma primeira descrição baseada em corpus. In *Domínios de Linguagem* 5(1): 1-16.
http://www.seer.ufu.br/index.php/dominiosdelinguagem/article/view/12479

Martin, P. 2005. WinPitch Corpus: a text-to-speech analysis ans alignment tool. In Cresti, E. & Moneglia, M. (eds), *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: John Benjamins, 40-51.

Mello, H., Raso, T., Mittmann, M., Vale, H. & Côrtes, P. In Press. Transcrição e segmentação prosódica do *corpus* C-ORAL-BRASIL: critérios de implementação e validação. In T. Raso & H. Mello (eds) In Press.

Mello, H. & Raso, T. 2009. Para a transcriçao da fala espontânea: o caso do C-ORAL-BRASIL. *Revista Portuguesa de Humanidades* 13(1): 153-178.

Moneglia, M. 2011. Spoken Corpora and Pragmatics. *Revista Brasileira de Linguística Aplicada* 11(2): 479-519. http://www.periodicos.letras.ufmg.br/rbla/arquivos/335.pdf

Moneglia, M. & Cresti, E. 1997. L'intonazione e i criteri di trascrizione del parlato adulto e infantile. In U. Bortolini & E. Pizzuto (eds), *Il Progetto CHILDES Italia*. Pisa: Del Cerro, 57-90.

Moneglia, M. & Cresti, E. 2006. C-ORAL-ROM. Prosodic boundaries for spontaneous speech analysis. In Y. Kawaguchi, S. Zaima & T. Takagaki (eds), *Spoken Language Corpus and Linguistics Informatics.* Amsterdam: John Benjamins, 89-112.

Panunzi, A. & Gregori, L. 2011. IPIC: an XML data base for the study of Informational Patterning. Presented at 7th LABLITA International Workshop in Corpus Linguistics, Firenze, http://lablita.dit.unifi.it/lablita_workshop/ws7

Raso, T. In Press. O C-ORAL-BRASIL e a Teoria da Língua em Ato. In T. Raso & H. Mello (eds). In Press.

Raso, T. & Leite, F. 2010. Estudo contrastivo do uso de Alocutivos em italiano, português e espanhol europeus e português brasileiro. *Domínios de Lingu@gem* 4(1): 151-174. http://www.dominiosdelinguagem.org.br/pdf/dl7/DL7-10.pdf

Raso, T. & Ulisses, A. 2008. Tópico e Apêndice no português do Brasil: algumas considerações. *Revista de Estudos da Linguagem* 16(1): 247-262. http://relin.letras.ufmg.br/revista/upload/11-Tommaso_Raso.pdf

Raso, T. & Mello, H. 2009. Parâmetros de compilação de um corpus oral: o caso do C-ORAL-BRASIL. *Veredas*: 20-35.

Raso, T. & Mello, H. 2010. The C-ORAL-BRASIL corpus. In M. Moneglia & A. Panunzi (eds), *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*. Università degli studi di Firenze, 193-213.

Raso, T. & Mello, H. in press. *C-ORAL-BRASIL I. Corpus de referência para a fala espontânea informal do português do Brasil*. Belo Horizonte: UFMG.

Raso, T. & Mittmann, M. 2009. Validação estatística dos critérios de segmentação da fala espontânea no corpus C-ORAL-BRASIL. *Revista de Estudos da Linguagem* 17(2): 73-91. http://relin.letras.ufmg.br/revista/upload/17-2_04.pdf

Raso, T., Moraes, J. & Mittmann, M. (forthcoming). *A Topic Prosodic Form in Brazilian Portuguese*.

R Development Core Team. 2010. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. http://www.r-project.org

Scarano, A. 2009. A The prosodic annotation of C-ORAL-ROM and the structure of information in spoken language. In L. Mereu (ed.), *Information structures and its interfaces*. Berlin and New York: Mouton de Gruyter, 51-74.

Signorini, S. 2003. Il Topic: criteri di identificazione e correlati morfosintattici in un corpus di italiano parlato. In F. Albano Leoni, F. Cutugno, M. Pettorino & R. Savy (eds), *Il parlato*

italiano. Atti del Convegno Nazionale. Firenze: Franco Cesati, 227-238. http://lablita.dit.unifi.it/preprint/preprint-03coll05.pdf

Tucci, E. 2006. L'unità di appendice in un corpus di italiano parlato (C-ORAL-ROM): caracteristiche intonative, semantiche e morfo-sintattiche. BA thesis, Facoltà di lettere e filosofia, Universitá degli studi di Firenze.

Tucci, I. 2004. L'inciso: caratteristiche morfosintattiche e intonative in un corpus di riferimento. In F. Albano Leoni, F. Cutugno, M. Pettorino & R. Savy (eds), *Il parlato italiano. Atti del Convegno Nazionale*. Napoli: D'Auria Editore, 1-14.

Tucci, I. 2009. "Obiter dictum": la funzione informativa delle unità parentetiche. In *Atti del Convegno Internazionale G.S.C.P. "La comunicazione parlata"*. Napoli.

Ulisses, A. 2008. A unidade de Apêndice no português do Brasil. BA thesis, Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte.

STRUMENTI
PER LA DIDATTICA E LA RICERCA

1. Brunetto Chiarelli, Renzo Bigazzi, Luca Sineo (a cura di), *Alia: Antropologia di una comunità dell'entroterra siciliano*
2. Vincenzo Cavaliere, Dario Rosini, *Da amministratore a manager. Il dirigente pubblico nella gestione del personale: esperienze a confronto*
3. Carlo Biagini, *Information technology ed automazione del progetto*
4. Cosimo Chiarelli, Walter Pasini (a cura di), *Paolo Mantegazza. Medico, antropologo, viaggiatore*
5. Luca Solari, *Topics in Fluvial and Lagoon Morphodynamics*
6. Salvatore Cesario, Chiara Fredianelli, Alessandro Remorini, *Un pacchetto evidence based di tecniche cognitivo-comportamentali sui generis*
7. Marco Masseti, *Uomini e (non solo) topi. Gli animali domestici e la fauna antropocora*
8. Simone Margherini (a cura di), *BIL Bibliografia Informatizzata Leopardiana 1815-1999: manuale d'uso ver. 1.0*
9. Paolo Puma, *Disegno dell'architettura. Appunti per la didattica*
10. Antonio Calvani (a cura di), *Innovazione tecnologica e cambiamento dell'università. Verso l'università virtuale*
11. Leonardo Casini, Enrico Marone, Silvio Menghini, *La riforma della Politica Agricola Comunitaria e la filiera olivicolo-olearia italiana*
12. Salvatore Cesario, *L'ultima a dover morire è la speranza. Tentativi di narrativa autobiografica e di "autobiografia assistita"*
13. Alessandro Bertirotti, *L'uomo, il suono e la musica*
14. Maria Antonietta Rovida, *Palazzi senesi tra '600 e '700. Modelli abitativi e architettura tra tradizione e innovazione*
15. Simone Guercini, Roberto Piovan, *Schemi di negoziato e tecniche di comunicazione per il tessile e abbigliamento*
16. Antonio Calvani, *Technological innovation and change in the university. Moving towards the Virtual University*
17. Paolo Emilio Pecorella, *Tell Barri/Kahat: la campagna del 2000. Relazione preliminare*
18. Marta Chevanne, *Appunti di Patologia Generale. Corso di laurea in Tecniche di Radiologia Medica per Immagini e Radioterapia*
19. Paolo Ventura, *Città e stazione ferroviaria*
20. Nicola Spinosi, *Critica sociale e individuazione*
21. Roberto Ventura (a cura di), *Dalla misurazione dei servizi alla customer satisfaction*
22. Dimitra Babalis (a cura di), *Ecological Design for an Effective Urban Regeneration*
23. Massimo Papini, Debora Tringali (a cura di), *Il pupazzo di garza. L'esperienza della malattia potenzialmente mortale nei bambini e negli adolescenti*
24. Manlio Marchetta, *La progettazione della città portuale. Sperimentazioni didattiche per una nuova Livorno*
25. Fabrizio F.V. Arrigoni, *Note su progetto e metropoli*
26. Leonardo Casini, Enrico Marone, Silvio Menghini, *OCM seminativi: tendenze evolutive e assetto territoriale*
27. Pecorella Paolo Emilio, Raffaella Pierobon Benoit, *Tell Barri/Kahat: la campagna del 2001. Relazione preliminare*
28. Nicola Spinosi, *Wir Kinder. La questione del potere nelle relazioni adulti/bambini*
29. Stefano Cordero di Montezemolo, *I profili finanziari delle società vinicole*
30. Luca Bagnoli, Maurizio Catalano, *Il bilancio sociale degli enti non profit: esperienze toscane*
31. Elena Rotelli, *Il capitolo della cattedrale di Firenze dalle origini al XV secolo*
32. Leonardo Trisciuzzi, *Barbara Sandrucci, Tamara Zappaterra, Il recupero del sé attraverso l'autobiografia*
33. Nicola Spinosi, *Invito alla psicologia sociale*
34. Raffaele Moschillo, *Laboratorio di disegno. Esercitazioni guidate al disegno di arredo*
35. Niccolò Bellanca, *Le emergenze umanitarie complesse. Un'introduzione*

36. Giovanni Allegretti, *Porto Alegre una biografia territoriale. Ricercando la qualità urbana a partire dal patrimonio sociale*
37. Riccardo Passeri, Leonardo Quagliotti, Christian Simoni, *Procedure concorsuali e governo dell'impresa artigiana in Toscana*
38. Nicola Spinosi, *Un soffitto viola. Psicoterapia, formazione, autobiografia*
39. Tommaso Urso, *Una biblioteca in divenire. La biblioteca della Facoltà di Lettere dalla penna all'elaboratore. Seconda edizione rivista e accresciuta*
40. Paolo Emilio Pecorella, Raffaella Pierobon Benoit, *Tell Barri/Kahat: la campagna del 2002. Relazione preliminare*
41. Antonio Pellicanò, *Da Galileo Galilei a Cosimo Noferi: verso una nuova scienza. Un inedito trattato galileiano di architettura nella Firenze del 1650*
42. Aldo Burresi (a cura di), *Il marketing della moda. Temi emergenti nel tessile-abbigliamento*
43. Curzio Cipriani, *Appunti di museologia naturalistica*
44. Fabrizio F.V. Arrigoni, *Incipit. Esercizi di composizione architettonica*
45. Roberta Gentile, Stefano Mancuso, Silvia Martelli, Simona Rizzitelli, *Il Giardino di Villa Corsini a Mezzomonte. Descrizione dello stato di fatto e proposta di restauro conservativo*
46. Arnaldo Nesti, Alba Scarpellini (a cura di), *Mondo democristiano, mondo cattolico nel secondo Novecento italiano*
47. Stefano Alessandri, *Sintesi e discussioni su temi di chimica generale*
48. Gianni Galeota (a cura di), *Traslocare, riaggregare, rifondare. Il caso della Biblioteca di Scienze Sociali dell'Università di Firenze*
49. Gianni Cavallina, *Nuove città antichi segni. Tre esperienze didattiche*
50. Bruno Zanoni, *Tecnologia alimentare 1. La classe delle operazioni unitarie di disidratazione per la conservazione dei prodotti alimentari*
51. Gianfranco Martiello, *La tutela penale del capitale sociale nelle società per azioni*
52. Salvatore Cingari (a cura di), *Cultura democratica e istituzioni rappresentative. Due esempi a confronto: Italia e Romania*
53. Laura Leonardi (a cura di), *Il distretto delle donne*
54. Cristina Delogu (a cura di), *Tecnologia per il web learning. Realtà e scenari*
55. Luca Bagnoli (a cura di), *La lettura dei bilanci delle Organizzazioni di Volontariato toscane nel biennio 2004-2005*
56. Lorenzo Grifone Baglioni (a cura di), *Una generazione che cambia. Civismo, solidarietà e nuove incertezze dei giovani della provincia di Firenze*
57. Monica Bolognesi, Laura Donati, Gabriella Granatiero, *Acque e territorio. Progetti e regole per la qualità dell'abitare*
58. Carlo Natali, Daniela Poli (a cura di), *Città e territori da vivere oggi e domani. Il contributo scientifico delle tesi di laurea*
59. Riccardo Passeri, *Valutazioni imprenditoriali per la successione nell'impresa familiare*
60. Brunetto Chiarelli, Alberto Simonetta, *Storia dei musei naturalistici fiorentini*
61. Gianfranco Bettin Lattes, Marco Bontempi (a cura di), *Generazione Erasmus? L'identità europea tra vissuto e istituzioni*
62. Paolo Emilio Pecorella, Raffaella Pierobon Benoit, *Tell Barri / Kahat. La campagna del 2003*
63. Fabrizio F.V. Arrigoni, *Il cervello delle passioni. Dieci tesi di Adolfo Natalini*
64. Saverio Pisaniello, *Esistenza minima. Stanze, spazî della mente, reliquiario*
65. Maria Antonietta Rovida (a cura di), *Fonti per la storia dell'architettura, della città, del territorio*
66. Ornella De Zordo, *Saggi di anglistica e americanistica. Temi e prospettive di ricerca*
67. Chiara Favilli, Maria Paola Monaco, *Materiali per lo studio del diritto antidiscriminatorio*
68. Paolo Emilio Pecorella, Raffaella Pierobon Benoit, *Tell Barri / Kahat. La campagna del 2004*
69. Emanuela Caldognetto Magno, Federica Cavicchio, *Aspetti emotivi e relazionali nell'e-learning*
70. Marco Masseti, *Uomini e (non solo) topi* (2ª edizione)
71. Giovanni Nerli, Marco Pierini, *Costruzione di macchine*
72. Lorenzo Viviani, *L'Europa dei partiti. Per una sociologia dei partiti politici nel processo di integrazione europea*
73 Teresa Crespellani, *Terremoto e ricerca. Un percorso scientifico condiviso per la caratterizzazione del comportamento sismico di alcuni depositi italiani*

74 Fabrizio F.V. Arrigoni, *Cava. Architettura in "ars marmoris"*
75. Ernesto Tavoletti, *Higher Education and Local Economic Development*
76. Carmelo Calabrò, *Liberalismo, democrazia, socialismo. L'itinerario di Carlo Rosselli (1917-1930)*
77. Luca Bagnoli, Massimo Cini (a cura di), *La cooperazione sociale nell'area metropolitana fiorentina. Una lettura dei bilanci d'esercizio delle cooperative sociali di Firenze, Pistoia e Prato nel quadriennio 2004-2007*
78. Lamberto Ippolito, *La villa del Novecento*
79. Cosimo Di Bari, *A passo di critica. Il modello di* Media Education *nell'opera di Umberto Eco*
80. Leonardo Chiesi (a cura di), *Identità sociale e territorio. Il Montalbano*
81. Piero Degl'Innocenti, *Cinquant'anni, cento chiese. L'edilizia di culto nelle diocesi di Firenze, Prato e Fiesole (1946-2000)*
82. Giancarlo Paba, Anna Lisa Pecoriello, Camilla Perrone, Francesca Rispoli, *Partecipazione in Toscana: interpretazioni e racconti*
83. Alberto Magnaghi, Sara Giacomozzi (a cura di), *Un fiume per il territorio. Indirizzi progettuali per il parco fluviale del Valdarno empolese*
84. Dino Costantini (a cura di), *Multiculturalismo alla francese?*
85. Alessandro Viviani (a cura di), *Firms and System Competitiveness in Italy*
86. Paolo Fabiani, *The Philosophy of the Imagination in Vico and Malebranche*
87. Carmelo Calabrò, *Liberalismo, democrazia, socialismo. L'itinerario di Carlo Rosselli*
88. David Fanfani (a cura di), *Pianificare tra città e campagna. Scenari, attori e progetti di nuova ruralità per il territorio di Prato*
89. Massimo Papini (a cura di), *L'ultima cura. I vissuti degli operatori in due reparti di oncologia pediatrica*
90. Raffaella Cerica, *Cultura Organizzativa e Performance economico-finanziarie*
91. Alessandra Lorini, Duccio Basosi (a cura di), *Cuba in the World, the World in Cuba*
92. Marco Goldoni, *La dottrina costituzionale di Sieyès*
93. Francesca Di Donato, *La scienza e la rete. L'uso pubblico della ragione nell'età del Web*
94. Serena Vicari Haddock, Marianna D'Ovidio, *Brand-building: the creative city. A critical look at current concepts and practices*
95. Ornella De Zordo (a cura di), *Saggi di Anglistica e Americanistica. Ricerche in corso*
96. Massimo Moneglia, Alessandro Panunzi (edited by), *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*
97. Alessandro Panunzi, *La variazione semantica del verbo* essere *nell'Italiano parlato*
98. Matteo Gerlini, *Sansone e la Guerra fredda. La capacità nucleare israeliana fra le due superpotenze (1953-1963)*
99. Luca Raffini, *La democrazia in mutamento: dallo Stato-nazione all'Europa*
100. Gianfranco Bandini (a cura di), *noi-loro. Storia e attualità della relazione educativa fra adulti e bambini*
101. Anna Taglioli, *Il mondo degli altri. Territori e orizzonti sociologici del cosmopolitismo*
102. Gianni Angelucci, Luisa Vierucci (a cura di), *Il diritto internazionale umanitario e la guerra aerea. Scritti scelti*
103. Giulia Mascagni, *Salute e disuguaglianze in Europa*
104. Elisabetta Cioni, Alberto Marinelli (a cura di), *Le reti della comunicazione politica. Tra televisioni e social network*
105. Cosimo Chiarelli, Walter Pasini (a cura di), *Paolo Mantegazza e l'Evoluzionismo in Italia*
106. Andrea Simoncini (a cura di), *La semplificazione in Toscana. La legge n. 40 del 2009*
107. Claudio Borri, Claudio Mannini (edited by), *Aeroelastic phenomena and pedestrian-structure dynamic interaction on non-conventional bridges and footbridges*
108. Emiliano Scampoli, *Firenze, archeologia di una città (secoli I a.C. – XIII d.C.)*
109. Emanuela Cresti, Iørn Korzen (a cura di), *Language, Cognition and Identity. Extensions of the endocentric/exocentric language typology*
110. Alberto Parola, Maria Ranieri, *Media Education in Action. A Research Study in Six European Countries*
111. Lorenzo Grifone Baglioni (a cura di), *Scegliere di partecipare. L'impegno dei giovani della provincia di Firenze nelle arene deliberative e nei partiti*

112. Alfonso Lagi, Ranuccio Nuti, Stefano Taddei, *Raccontaci l'ipertensione. Indagine a distanza in Toscana*
113. Lorenzo De Sio, *I partiti cambiano, i valori restano? Una ricerca quantitativa e qualitativa sulla cultura politica in Toscana*
114. Anna Romiti, *Coreografie di stakeholders nel management del turismo sportivo*
115. Guidi Vannini (a cura di), *Archeologia Pubblica in Toscana: un progetto e una proposta*
116. Lucia Varra (a cura di), *Le case per ferie: valori, funzioni e processi per un servizio differenziato e di qualità*
117. Gianfranco Bandini (a cura di), *Manuali, sussidi e didattica della geografia. Una prospettiva storica*
118. Anna Margherita Jasink, Grazia Tucci e Luca Bombardieri (a cura di), *MUSINT. Le Collezioni archeologiche egee e cipriote in Toscana. Ricerche ed esperienze di museologia interattiva*
119. Ilaria Caloi, *Modernità Minoica. L'Arte Egea e l'Art Nouveau: il Caso di Mariano Fortuny y Madrazo*
120. Heliana Mello, Alessandro Panunzi, Tommaso Raso (edited by), *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation*