UNIVERSIDADE FEDERAL DE MINAS GERAIS
Faculdade de Letras
Programa de Pós-graduação em Estudos Linguísticos

Frederico Amorim Cavalcante

THE INFORMATION UNIT OF TOPIC:
a crosslinguistic, statistical study based on spontaneous speech corpora

Belo Horizonte
2020

Frederico Amorim Cavalcante

THE INFORMATION UNIT OF TOPIC:
a crosslinguistic, statistical study based on spontaneous speech corpora

Tese apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Doutor em Linguística teórica e descritiva.

Área de Concentração: Linguística teórica e descritiva
Linha de Pesquisa: Estudos linguísticos baseados em corpora
Orientador: Prof. Dr. Tommaso Raso
Coorientador: Prof. Dr. Plínio A. Barbosa

Belo Horizonte
2020

Ficha catalográfica elaborada pelos Bibliotecários da Biblioteca FALE/UFMG

C376i Cavalcante, Frederico Amorim.
The Information Unit of Topic [manuscrito] : a crosslinguistic, statistical study based on spontaneous speech corpora / Frederico Amorim Cavalcante. – 2020.
190 f., enc. : il., tabs., p&b.

Orientador: Tommaso Raso.

Coorientador: Plínio A. Barbosa.

Área de concentração: Linguística Teórica e Descritiva.

Linha de pesquisa: Estudos Linguísticos Baseados em Corpora.

Tese (doutorado) – Universidade Federal de Minas Gerais, Faculdade de Letras.

Bibliografia: f. 169-179.

Anexos: f. 180-190.

Inclui CD com anexos utilizados na pesquisa.

1. Linguística de corpus – Teses. 2. Atos de fala (Linguística) – Teses. 3. Análise linguística – Teses. I. Raso, Tommaso. II. Barbosa, Plínio A. (Plínio Almeida), 1966- III. Universidade Federal de Minas Gerais. Faculdade de Letras. IV. Título.

CDD : 414

UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGUÍSTICOS

PosLin

FOLHA DE APROVAÇÃO

The information unit of topic: a crosslinguistic, statistical study based on spontaneous speech corpora

FREDERICO AMORIM CAVALCANTE

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTUDOS LINGUÍSTICOS, como requisito para obtenção do grau de Doutor em ESTUDOS LINGUÍSTICOS, área de concentração LINGUÍSTICA TEÓRICA E DESCRITIVA, linha de pesquisa Estudos Linguísticos Baseados em Corpora.

Aprovada em 11 de fevereiro de 2020, pela banca constituída pelos membros:

Prof(a). Tommaso Raso - Orientador
Universidade Federal de Minas Gerais

Prof(a). Plinio Almeida Barbosa (Coorientador)
UNICAMP

Prof(a). Heliana Ribeiro de Mello
UFMG

Prof(a). Maria Mendes Cantoni
UFMG

Prof(a). Albert Olivier Blaise Rilliard
Limsi, Paris

Prof(a). Donna Erickson
Kanazawa Medical University

Belo Horizonte, 11 de fevereiro de 2020.

Profa. Ana Larissa Adorno Marciotto Oliveira
Subcoord. Programa de Pós- Graduação
em Estudos Linguísticos
FALE/UFMG

À vó Socorro <3

Agradecimentos

Ao Prof. Tommaso Raso
Ao Prof. Plínio A. Barbosa
À Banca Examinadora
À Capes
Aos colegas do LEEL
Aos amigos
À Profa. Heliana Mello
À minha mãe
Ao meu pai
À Lúcia, Gilda e Pequi
Ao Douglas

Resumo

Realizada nos quadros da *Language into Act Theory* (L-AcT), esta tese constitui um estudo estatístico e interlinguístico da unidade informacional de Tópico (TOP), cuja função é fornecer um domínio cognitivo para a interpretação do Comentário (COM), ou seja, a unidade informacional que veicula a força ilocucionária. Foram examinados dados de corpora de fala espontânea do italiano (IT), do português europeu (PE) e brasileiro (PB) e do inglês americano (IA), todos compilados segundo os princípios da L-AcT, contendo, portanto, variedade de situações comunicativas, segmentação prosódica e alinhamento texto-som ao nível do enunciado. Alguns desses corpora também possuem anotação informacional, especificamente os minicorpora do IT, PB e IA. Destes, os dois últimos passaram por uma revisão como parte da preparação dos dados pare este estudo. Realizou-se uma discussão das abordagens mais relevantes sobre *estrutura informacional* em geral, dedicando atenção especial ao termo *tópico* e as diferentes acepções que apresenta dependendo do quadro teórico considerado. Um teste kappa foi conduzido para estabelecer o grau de acordo entre anotadores numa tarefa de identificação de TOP utilizando-se dados dos corpora C-ORAL-BRASIL II. Fez-se uso da *Análise de Dados Funcionais* (FDA, da sigla em inglês) e da *Análise de Componentes Principais Funcionais* (F-PCA, idem) para verificar se as formas prosódicas de TOP propostas em estudos anteriores são separáveis bem como para propor modelos estatísticos para cada uma das formas. Além disso, utilizou-se a *Análise de Variância* (ANOVA) para verificar se as durações das sílabas nucleares e não nucleares de TOP são significativamente diferentes. Os resultados deste estudo mostram que o acordo entre anotadores na identificação de TOP é substancial. Além disso, os resultados fornecem evidências estatísticas convincentes que corroboram o esquema classificatório de estudos anteriores, os quais não haviam sido validados estatisticamente. O estudo da duração, por sua vez, mostrou que as sílabas nucleares de TOP são significativamente mais longas do que as sílabas não nucleares. Finalmente, as curvas melódicas de TOP foram comparadas a curvas de Comentários Ligados (COB), os quais constituem unidades ilocucionárias que exibem sinal prosódico de continuidade. Isso também foi realizado por meio das técnicas de FDA e F-PCA, com o objetivo de mostrar que o TOP não é meramente uma unidade portadora de sinal de continuidade. A análise mostra que as curvas de f0 de TOP são separáveis das de COB e que, nos casos em que há similaridade formal, o número de sílabas nucleares e a posição dessas sílabas na unidade distinguem claramente o TOP do COB.

Palavras-chave: Language into Act Theory. Estrutura informacional. Tópico. Análise estatística.

Abstract

This study, carried out within the framework of the *Language into Act Theory* (L-AcT), constitutes a statistically based, crosslinguistic analysis of the information unit of Topic (TOP), whose function is to supply a cognitive domain for the interpretation of the Comment (COM), the information unit that conveys the illocutionary force. We looked at data from spontaneous speech corpora of Italian (IT), European Portuguese (EP), Brazilian Portuguese (BP), and American English (AE), all of which collected in accordance with the tenets and methods established by L-AcT, thus recording a variety of formal and informal communicative situations, prosodic segmentation, and text-to-speech alignment at the utterance level. Some of these linguistic resources also feature annotation of informational functions, particularly the IT, BP and AE minicorpora. The last two of these have been revised as part of the preparation of the data for this study. We conducted a critical review of some of the most prominent approaches to *information structure* in general, dedicating more attention to the term *topic* and the different meanings it has depending on the framework considered. We conducted a kappa test to establish the degree of agreement among four raters in a task of TOP annotation based on data from the C-ORAL-BRASIL II corpora. We used *Functional Data Analysis* (FDA) and *Functional Principal Component Analysis* (F-PCA) to verify whether the prosodic forms of TOP proposed in previous studies are separable and to propose statistical models for each of the forms. In addition, we used *Analysis of Variance* (ANOVA) to determine whether the nuclear and non-nuclear syllables of TOP are significantly different from each other in terms of duration. Our study has shown that the interrater agreement for the detection of TOP in spontaneous speech is substantial. Furthermore, it has provided compelling statistical evidence in support of the classification scheme for the prosodic forms of TOP proposed in previous studies, which are lacking in statistical validation. In addition, the study of syllable durations has shown that the nuclear syllables of TOP are significantly longer than the non-nuclear ones. Finally, the melodic curves of TOP were compared to those of Bound Comments (COB), which constitute illocutionary units that feature a prosodic signal of continuity. This was done using the FDA and F-PCA techniques with the aim to show that TOP is not simply a simply a bearer of a continuity signal. The analysis showed that the f0 curves of COB are separable from those of TOP, and in those cases where there is some similarity in the shape of TOP and COB curves, the number of nuclear syllables as well as their position within the unit clearly distinguish these two types of information unit.

Keywords: Language into Act Theory. Information structure. Topic. Statistical analysis.

List of Figures

List of Tables

List of abbreviations

AE	American English
AdjP	Adjectival phrase
AdvP	Adverbial phrase
ALL	Allocutive
ANOVA	Analysis of Variance
APA	Appendix
APC	Appendix of Comment
APT	Appendix of Topic
AUX	Dialogic unit with unspecified subfunction
BP/EP	Brazilian/European Portuguese
CMM	Multiple Comment
COB	Bound Comment
COM	Comment
CNT	Conative
DCT	Discourse Connector
EMP	Empty unit
EP	European Portuguese
EXP	Expressive
f0	Fundamental frequency
FDA	Functional Data Analysis
F-PCA	Functional Principal Component Analysis
INT	Locutive Introducer
INP	Incipit
IS	Information structure
i-[TAG]	Interrupted [information unit]
L-AcT	Language into Act Theory
NP	Noun phrase
PAR	Parenthetic
PHA	Phatic
SCA	Scanning unit
TMT	Time taking unit (filled pause)
TOP	Topic
TPL	List of topics
[TAG]_r	Information unit in Reported Speech
UNC	Unidentifiable information unit
VP	Verb phrase

List of symbols

*	Beginning of dialogic turn
[/n]	Tone unit boundary with retracting
/	Non-terminal prosodic break
//	Terminal prosodic break
+	Utterance interruption
<	Beginning of overlapping speech
>	End of overlapping speech
&	Beginning of interrupted word or hesitation
&he	Filled pauses
%ill	Illocutionary value
xxx	One incomprehensible word
yyy	More than one incomprehensible word
hhh	Paralinguistic noise
yyyy	Anonymized information

Contents

Introduction

This dissertation constitutes a statistic and crosslinguistic study on information structure which was carried out within the framework of the *Language into Act Theory* (L-AcT), a corpus-based approach for the study of speech (CRESTI, 2000, MONEGLIA; RAS0, 2014, CAVALCANTE, 2015). Stemming from Austin's (1962) *Speech Act Theory*, L-AcT regards the illocution – the "action dimension" of spoken language encoded prosodically – as a core aspect of speech behavior and the point of departure for the structuring of information. This study deals specifically with the information unit of *topic* (TOP), a unit that has the function of supplying a cognitive domain for the interpretation of the illocution, which is carried by the *comment* (COM).

Statistical techniques were used in order to analyze speech data collected in corpora of three different languages, namely American English (AE), Brazilian and European Portuguese (BP and EP), and Italian (IT). The collection of corpora used in this study comprise:

- the AE minicorpus (CAVALCANTE; RAMOS, 2016)
- the BP minicorpus (RASO; MITTMANN, 2011, RASO; CRESTI, 2012)
- the EP component of the C-ORAL-ROM (NASCIMENTO et al., 2005, ROCHA, 2012)
- the IT minicorpus (RASO; CRESTI, 2012, RASO; BOSSAGLIA, forthcoming)

These corpora were compiled according to the tenets and methodological standards of L-AcT and are part of the C-ORAL family of corpora (CRESTI; MONEGLIA, 2005, RASO; MELLO, 2012). The main statistical techniques used were *Functional Data Analysis* (FDA) and *Functional Principal Component Analysis* (F-PCA; RAMSAY; SILVERMAN, 2005), *Analysis of Variance* (ANOVA), and the *Kappa statistic* (COHEN, 1960; FLEISS, 1971).

In deciding to study TOP within the L-AcT framework, this dissertation follows a number of other studies that have looked at the same information unit in IT (FIRENZUOLI; SIGNORINI, 2003, SIGNORINI, 2005), BP (MITTMANN, 2012), EP (ROCHA, 2012), and AE (CAVALCANTE, 2015, RASO et al. 2017), all of which conducted within L-AcT, but none validated statistically.

Goals and objectives

The present study aims to refine the description of TOP as understood within the L-AcT framework and, secondarily, to compare the way information structure is regarded within L-AcT – particularly the notion of topic[1] – to the way it is regarded within some other major linguistic frameworks. In order to do that we did (i) a literature review on information structure and topic, (ii) an interrater agreement test of the detection of TOP in spontaneous speech data, and (iii) an in-depth crosslinguistic study of the prosodic forms of TOP according to L-AcT.

The objectives of this study are as follows:

- to conduct a critical literature review contemplating information structure in general and the notions of topic across different approaches
- to conduct a test using the kappa statistic to measure the interrater agreement on the detection of TOP in spontaneous speech using data from C-ORAL-BRASIL II (RASO et. al, forthcoming)
- to revise the prosodic and informational annotation of the AE and BP minicorpora
- to use FDA and F-PCA to examine the melodic shapes of TOP in AE, BP, EP, and IT
- to generate statistical models for the melodic curves of the prosodic forms of TOP
- to assess the classification of the prosodic forms of TOP based on the numerical descriptors computed with F-PCA, and
- to analyze the durational patterns of TOP using the *SGDetector* script (BARBOSA, 2006) and ANOVA.

Organization of the text

Chapters 1 to 3 concentrate on the theoretical aspects related to the subject of this dissertation. In chapter 1, some of the major approaches to information structure in the linguistics literature are critically reviewed, thus providing a background to situate L-AcT. This framework is presented in chapter 2 using examples taken from the AE minicorpus. The audio files corresponding to the examples given throughout this dissertation are available in the

[1] Throughout this dissertation, "topic" will be used as a generic term referring to any notion associated with topic outside the L-AcT framework. The acronym TOP will be used specifically to refer to the notion of topic defined by L-AcT.

accompanying folder. In chapter 3, various notions associated with the term topic are discussed and compared to the notion advanced by L-AcT.

Chapter 4 is dedicated to the methods and resources that were used in this study. Chapter 5 contains the kappa test that was conducted to measure the interrater agreement on the detection of TOP. Chapter 6 presents the statistical analysis of the melodic shapes of TOP using FDA and F-PCA, while chapter 7 deals with the study of the duration patterns of the unit. Finally, chapter 8 summarizes the conclusions arrived at by means of the analysis carried out for this dissertation.

1 Information Structure

In this dissertation we delve into an aspect of information structure – namely, the information unit of Topic (TOP) – following the tenets of the *Language into Act Theory*[2] (L-AcT), which constitutes a framework for the study of spoken language that regards the speech act (AUSTIN, 1962) as the core facet of linguistic behavior and prosody as a privileged formal signal not only of the accomplishment of speech acts but also of the way information is organized in speech. In this chapter, we provide an overview of some of the major approaches to the study of information structure, thus paving the way for a critical consideration of L-AcT, which will be reviewed in the next chapter.

Information structure (IS) is an umbrella term that includes notions meant to account for linguistic phenomena that are constrained by – or sensitive to – context pressure. For instance, it is presumed that what the speaker assumes her addressee is conscious of at the time of the utterance accounts – at least in part – for the way that referents are linguistically encoded – e.g. an indexical form or a description – and the position that these linguistically encoded referents will occur in the utterance. In a similar vein, concepts such as *new* and *old* – or variants thereof – are often correlated with certain syntactic functions and phonological features. This chapter discusses IS and some of the main notions and terms that are relevant within frameworks of different theoretical persuasions. But before that, a short review is provided of the term *information* and some ideas related to it.

1.1 A few words about the term information

Information is a term that carries different meanings and receives different interpretations even within the field of *Information Theory* itself (SHANNON; WEAVER, 1949). In ordinary language, to be sure, the situation is not different. As Adriaans and Benthem (2008, p. 7) remind us, information is a very frequent word that has low semantic content and that is often ambiguous, sometimes implying truth, sometimes not. For instance, when we hear that someone has *information* about a fact, we tend to understand that whatever it is that the term information is referring to, that thing is true. In a similar vein, the information in a flight schedule display at an airport must be true in order to be of any value to us. At the same time, there is no

[2] For more on L-AcT, see chapter 2 in this dissertation and Cresti (2000), Moneglia and Raso (2014), and Cavalcante (2015).

contradiction in qualifying information as “vague”, “insufficient”, or “imprecise”, which shows that information and truth need not go hand in hand.

Many scholars recognize that it is hard to provide a unified definition for information, a term that has been part of the history of western thought at least since Cicero (103-43 BCE) and Augustine (354-430 CE). These two thinkers used the Latin terms *informare* and *informatio* as translations for Greek philosophical concepts such as *eîdos* ‘essence’, *idea* ‘idea’, *týpos* ‘type’, *morphé* ‘form’, and *prólepsis* ‘representation’ (ADRIAANS, 2017). Those concepts, used in the context of ontological and epistemological inquiries by Plato (427-347 BCE), Aristotle (384-322 BCE) and others, suggest that information has been associated with abstractions and mathematical representation since classical antiquity, given that *form* – meaning both “essence of” and “condition for knowing” an object – was considered a structure that could be described mathematically (ADRIAANS, 2017). Nevertheless, mathematical formalizations of information in the form of full-fledged theories would have to wait until the 20th century to be developed (ADRIAANS, 2017).

Within the context of *Philosophy of Information*, Floridi (2016a) points out that asking what information is requires asking other similarly challenging questions regarding the nature of truth, knowledge, ethics, and meaning. The scholar regards information – its meanings, uses, applications, and different kinds – as a distributed network of connected concepts. Centered on semantics, his approach is epistemically oriented and focuses on the knowledge that information can provide about its reference. He clarifies that the centrality given to semantics is not a replacement for other aspects related to information but rather “a hermeneutical device that influences, interrelates, and helps to access other notions” (FLORIDI, 2016a, p. 3).

From a different perspective, the influential framework of the *Mathematical Theory of Communication* (MTC) sees information as a measure of the uncertainty reduction associated with the outcome of a random process (SHANNON, 1948; SHANNON; WEAVER, 1949; FLORIDI, 2010, 2016b). Developed as a way to deal with electrical engineering problems involving the limits of communication, data compression, and data transmission, MTC does not address information from a semantic standpoint, as it is not primarily concerned with those aspects related to meaning, reference, relevance, reliability, or truthfulness.

According to Floridi, the difference between information within MTC and semantic information “is comparable to the difference between a Newtonian description of the physical laws describing the dynamics of a car accident and the description of the same accident by the police” (FLORIDI, 2016b, p. 44). MTC focuses on aspects such as source, signal, noise, and

receiver, aiming to provide formal means to quantifying the reduction of uncertainty associated with the source given what is obtained by a receiver through a signal.

Turning now to aspects more closely related to the topic of this study, natural languages, from a certain perspective, can be thought of as vehicles of information. The ability of a language to carry information in so many ways about so many different things is suggestive of its efficiency, adaptability, and usefulness (KAMP; STOKHOF, 2008). One may question this idea of language as a vehicle of information on the basis of utterances which, instead of describing states of affairs, elicit information, provide instructions, express emotions, and so on. But, to be sure, non-factual utterances, too, can be approached from the point of view of information, after all "a question must convey what information is requested; a directive must specify information about what is to be done, or to be refrained from; a warning threat must identify a particular situation or event (...)" (KAMP; STOKHOF, 2008, p. 53).

Linguistic approaches to IS, more often than not, use the term information without defining it clearly. In the introduction to the 2016 edition of the *Oxford Handbook of Information Structure*, to give a recent example, Féry and Ishihara (2016) provide a historical outline of IS studies and definitions, exploring the most common concepts dealt with in those studies and explored in the chapters of the volume. Those concepts include *information packaging*, *common ground*, *focus*, *givenness*, *topic*, and a few others, but information itself is taken for granted, never being explicitly defined.

Lambrecht (1994) constitutes an exception to this trend, as the author devotes a full chapter to discussing information. In it, information is regarded as a matter of pragmatic construal or, as the author puts it, as "the establishment of a relation between terms in a proposition" (LAMBRECHT, 1994, p. 48). In this perspective, morphological or syntactic elements are not considered vehicles of information. To inform, in a formulation which Lambrecht admittedly draws from Dahl (1976), is to influence the hearer's mental representation of the world, i.e. to induce a change in the hearer's knowledge by altering or adding a mental representation assumed to be possessed by the hearer at the time immediately prior to the enunciation.

Lambrecht (1994) uses the term *proposition* to refer to the mental representations of events, situations, and states of affairs that make up one's knowledge of the world, a perspective that is different from more traditional approaches that regard propositions as the content of assertions and truth-value bearers. According to the author,

> [o]ne can know, or be ignorant of, a certain event denoted by a proposition, i.e. one may, or may not, have some mental *picture* of the event; and one may be thinking of

> the event, or be oblivious of it, at a certain time, i.e. one may, or may not, have that picture at the forefront of one's consciousness. But to characterize the event, or the picture of it, as true or false seems incongruous (LAMBRECHT, 1994, p. 45, emphasis in the original).

Lambrecht goes on to say that a mental representation that proves to be false would be preferably characterized as *outdated* rather than false. We will return to these ideas later in this chapter, but at this point let us just mention that, for this author, the *pragmatic states* of denotations – i.e. their being new or old in some sense – are distinguished from the *pragmatic relations* that are established among them as elements associated in a mental picture.

Since the establishment of those pragmatic relations is precisely what gives rise to information, individual sentence constituents cannot be said to convey any information according to this approach. But information for Lambrecht is still propositional and based on a model of cognition as a form of computation. We will see that, according to the framework that guides this dissertation, information is something different, something that is inextricably related to the actional dimension and elaborated from a perspective that regards cognition in terms of embodied representations (see CARUANA; BORGHI, 2016).

As said above, the term information has indeed so many facets that it cannot be exhausted in a few paragraphs. The purpose of this section is simply to introduce some of ideas that will serve the discussion carried in the remainder of this chapter. But we expect that, by examining how IS and the notions associated with it are dealt with across different frameworks, more details about information itself will come to light.

1.2 Information structure across frameworks

Some of the ideas that are currently considered under the heading of IS – e.g. those associated with the terms *theme/rheme*, *presupposition/focus*, *topic/comment*, and the like – have been around for a long time. Some of them may be traced back at least to Henri Weil (1818-1909), whose work is often said to forerun some notions that are currently subsumed under IS and that came to be investigated by many linguists from the turn of the 20th century onwards[3]. Around the middle of the 19th century, Weil, discussing the nature of discourse, writes that every utterance consists of two parts, one of which constituting

[3] See, among others, Ping (2004), Krifka (2008), Marogy (2010), Krifka and Musan (2012), Féry and Ishihara (2016).

> […] a point of departure, an initial notion which is equally present to him who speaks and to him who hears, which forms, as it were, the ground upon which the two intelligences meet, and another part of discourse which forms the statement (*l'énonciation*), properly so called. This division is found in almost all we say (WEIL, 1978 [1844], p. 29, apud PING, 2004, p. 23).

In the passage quoted above, we see the association of utterance structure and memory based on the cognitive status of what is being said, suggesting a discourse model in terms of a chain where the "unknown" is constantly being hooked to what is already present to both speaker and hearer.

The term IS itself, however, came about much later, being accredited to Halliday (1967b, p. 199), who regards information not simply in terms of "content" but rather as a product of the tension between what is new and what is old in discourse. According to Halliday, this tension can be noted in the way the speaker organizes what she says into tone groups, in the placement of prosodic prominences inside the tone group, and also in how constituents are combined to form clauses and other syntactic structures. It appears that Halliday is also one of the first scholars to establish an indissociable relation between IS and prosody.

With ideas developed based on the work of linguists of the Prague circle, Halliday (1967b) recognizes the independence of information system and sentence structure, as he sees no necessary correspondence between information status and formal elements, though he points to correlations. He alludes to the possible existence of a unit pertaining to a level to which the tone group is subordinated, suggesting the existence of something between the tone group and the text but saying very little about it, as the passage below shows.

> The distribution into information units represents the speaker's blocking out of the message into quanta of information, or message blocks. Each information unit is realized as one tone group, in the sense that the information structure specifies the boundaries of the tone group to within certain limits, its exact location being determined by considerations of *phonological structure*. It may be that, above the information unit, it is possible to recognize *a higher unit of information structure* realized in terms of patterns of tone group sequences specified by tone (HALLIDAY, 1967b, p. 202, emphasis added)

In Halliday's framework, no association is made between the prosodic features of the tone group and specific pragmatic functions besides the signaling of old and new. The framework is very much centered on morphosyntax, and, as will be shown below, notions like *theme* and *rheme* are defined without an appropriate consideration of prosodic aspects according to the perspective that guides this study.

As mentioned previously, Halliday's framework was developed based on earlier contributions made by linguists of the Prague Circle. Mathesius (1983 [1929]), followed by other scholars – see e.g. Firbas (1964, 1992), Vachek (1966) –, established the so-called

Functional Sentence Perspective (FSP), a framework within which the sentence is regarded from the point of view of the degree to which elements contribute to push the communication forward or not. The "communicative weight", so to speak, of a sentence element is referred to as its *communicative dynamism* (CD).

Aiming to provide a functional definition of *sentence*, Mathesius (1983 [1929]) notes that declarative sentences exhibit an *active element* – a formal expression of the attitude of the speaker towards the state of affairs denoted by the sentence – in the form of *assertiveness*. This active element can take the form of a *thetic* sentence – also called *one-part* sentence –, which consists of a simple act of presentation of an event, situation or action. In addition, this active element can take the form of a *predicative sentence*, which is structured around two acts: the presentation of an entity and a statement about it. As Mathesius later wrote about the utterance of predicative sentences,

> when observing different utterances we find that they are more or less clearly composed of two parts. One part expresses what is given by the context or what naturally presents itself, in short *what is being commented upon*. […] this part is called the *theme* of the utterance. The second part contains the new element of the utterance, i.e. *what is being stated about something*; this part is called the *rheme* of the utterance (MATHESIUS, 1975 [1961], p. 156, emphasis added).

Sasse (1987) points out that Mathesius's approach likely stems from Brentano (1874), who first proposed the thetic/categorical distinction[4]. This distinction is based on the idea that human *judgement* can take the form of either a simple recognition (i.e. a thetic judgement) or a recognition plus a predication (i.e. a double or categorical judgment). In the first case, states of affairs take, as a whole, a presentational form, thus the sentences that convey this type of judgment cannot be construed as being about one of its constituents. Typical examples of thetic judgments are sentences such as "it's raining" or "it's dark".

The categorical judgement, on the other hand, gives rise to two-part sentences consisting of the recognition of an entity and a statement about it, or as Mathesius would put it, a theme and a rheme. An example of categorical sentence would be "my neck hurts", uttered in response to a question like "what's the matter with your neck?" (see LAMBRECHT, 1987). This distinction is present – though sometimes indirectly and with particular connotations – in the work of linguists such as Bolinger (1954), Hockett (1958), Chafe (1964), Halliday (1967b), Kuroda (1972), Kuno (1972), and, more recently, Lambrecht (1994) and Krifka (2012), many

[4] See Mulligan (1990).

of which to be discussed below in the sections devoted to specific categories of information structure.

Chafe constitutes crucial references in the IS literature (CHAFE, 1974, 1976). His notion of *information packaging* is widely cited and explored by many authors. It describes the way the speaker presents a proposition in discourse, i.e. it applies to form rather than semantic content. According to Chafe (1976, p. 28), information packaging can be thought of as analogous to a product package, which may say little about the product itself, but which can indeed affect the sales. Moreover, information packaging constitutes formal signals that allegedly disclose the speaker's assumptions regarding the state of mind – including memory, attention, and beliefs – of the addressee at the time of the communicative act.

The notion of common ground (STALNAKER, 1970, 1974, 2014) has to be mentioned here, as it has played an important role in IS theorization. *Common ground* (CG) corresponds to "a body of information that is presumed to be shared by the parties to a discourse" (STALNAKER, 2014, p. 2). It is associated with the beliefs (propositional attitudes) of individuals taking part in an exchange and, if the speakers are correct in their assumptions as to CG content, it overlaps with *speakers' presuppositions*. These, in turn, are the propositions taken to be true and that forms part of the context in which the interaction takes place. CG is a very important notion for approaches to IS based on a model that considers the phenomenon of communication as a succession of propositions being exchanged and updated.

Specific concepts for which the notion of CG is relevant are Heim's interconnected notions of *file* and *file card* (HEIM, 1982), which are in a sense related to Karttunen's (1976) discourse referent (HEIM, 1982, p. 85). In Heim's approach, *file* is a metaphor for all the information that has been exchanged in a conversation at a given point, while *file card* is where each piece of information is stored and identified. Vallduvì's (1992, p. 54) *link*, a topic-like element restricted to the initial position of the sentence, is conceived of as an instruction to the hearer as to the file card to which the information provided in the utterance should be added. This perspective is also related to the concept of sentence topic by Reinhart (1981), which we will return to in chapter 3.

Vallduvì (1992) proposes that linguistic competence should have a component – which he terms *informatics* – responsible for a special type of meaning, one that cannot be accounted for by syntax and truth-conditional semantics. According to his theory, informatics is the component that allows for the generation and interpretation of information packaging understood in terms of a signal "with which the hearer is instructed by the speaker to retrieve

the information carried by the sentence and enter it into her/his knowledge-store" (VALLDUVÌ, 1992, p. 18). Information, for this author, is the part of the propositional content that is meant to contribute to the hearer's knowledge-store. Whatever is entered into the hearer's knowledge-store, Vallduvì identifies as information, saving the term knowledge for what was already present in the hearer's mind prior to the utterance.

Krifka proposes an interesting approach to IS that also incorporates the notion of CG (KRIFKA, 2006, 2008). Building on Chafe's (1976) *information packaging*, Krifka develops his idea of IS regarding communication as a constant change of CG. One can only utter a sentence like "I have a dog", for instance, if she does not consider the proposition "the speaker has a dog" as part of the CG. Hence the oddness of uttering "I had to take my dog to the vet, and I have a dog", which turns out to be inappropriate unless "I have a dog" is equivalent to "oh, I'd forgotten to tell you I have a dog" (see KRIFKA, 2008). In Krifka's perspective, the CG is made up not only of the propositions but also the entities introduced previously in the discourse. Such entities may be explicitly introduced by means of, say, an indefinite NP, accommodated – provided they are not too controversial –, or picked up by anaphoric expressions.

All these possibilities are captured under the notion of information packaging, as they are taken to be responses by the speaker to the temporary states of the mind of the addressee. Under certain circumstances, some formal elements usually regarded in terms of packaging phenomena may impact the semantic level of what is uttered by having truth condition effects. Let us consider the examples below.

(1.1)

a. What did John show Mary?
a'. John showed Mary the PICTURES.
b. Who did John show the pictures?
b'. John showed MARY the pictures.

c. John only showed MARY the pictures.
d. John only showed Mary the PICTURES.

(KRIFKA; MUSAN, 2012, p. 4)

According to the authors, the different prosodic prominences of (1.1a') and (1.1b') – uppercase letters denote pitch accent – can be interpreted solely in terms of information packaging, since the two utterances allegedly have the same truth conditions. They argue that those prominences signals that each utterance conform to a different CG, the first of which containing the information that John had shown Mary something, and the other containing the information that John had shown the pictures to someone. Under this interpretation, (1.1a') adds to the CG that the pictures constitute the thing that was shown, while (1.1b') adds that Mary is the one to whom the pictures were shown.

In this framework, when the use of a formal device does not affect the CG content, it is regarded as CG management. This is not the case with the utterances (1.1c) and (1.1d), which exhibit different truth conditions due to the interaction of the prominence and the particle "only". In these utterances, the same formal device – i.e. pitch accent – is being used to build different contents, hence its interpretation as a strategy for CG content. As the discussion above suggests, Krifka's approach to IS combines logic and pragmatics. In the first set of examples, we see what the author calls *pragmatic focus*, i.e. a focus that does not affect truth conditions; in the second, what he calls *semantic focus*, i.e. a focus that affects the truth conditions.

Advocating the incorporation of IS as a component of grammar, Lambrecht (1994, p. 1-6 and 120-122) argues that, while many morphosyntactic and prosodic characteristics of sentences may be analyzed without recourse to IS, only IS can account for the difference between sentences exhibiting alternative syntactic structures but identical propositional content, the so-called *allosentences*.

Lambrecht defines IS as

> that component of sentence grammar in which propositions *as conceptual representations of states of affairs* are paired with lexicogrammatical structures in accordance with the mental states of interlocutors who use and interpret these structures as units of information in given discourse contexts (LAMBRECHT, 1994, p. 5, emphasis added).

According to Lambrecht (1994, p. 46), information is what the speaker adds to the addressee's knowledge by means of an act of communication. Therefore, to inform is equivalent to cause a change in the mental representation that the addressee has of the world. The author makes sure to distinguish between sentence meaning, on the one hand, and utterance information, on the other. While *sentence meaning* is what remains constant regardless of pragmatic considerations, *utterance information* is dependent upon what the speaker considers to be shared with the addressee at the time of the utterance. In other words, it is the particular

context of a communicative act that will determine whether a piece of *propositional meaning*, to put it like Lambrecht, may be said to be informative or not.

Based on assumptions as to what constitutes the addressee's mental representation or picture of the world, the speaker makes her communicative move by choosing specific formal devices that signal the information status and the relation between discourse referents. In this perspective, the linguistic analysis of IS should not be concerned with the logical notion of truth. Lambrecht argues that mental pictures cannot be appropriately described in vericonditional terms – after all, a mental picture can be false and still be present in the mind. That is, the alleged presence and status of a representation in the mind of the addressee is what matters, and the fact that one can have a mental picture of a false proposition speaks to that: even in the face of an utterance like "it is not true that Sue is married", one can entertain the propositional content "Sue is married".

Within the framework of the *Inferential model* (see DOMANESCHI, 2016), to grasp the meaning of an utterance we must have access to the context in which it was enunciated. To communicate linguistically, in this perspective, is to use linguistic expressions as a clue to induce in the addressee a mental representation of the speaker's intention. For example, the expression "all of them" uttered as a response to a question such as "do you know any movies by Woody Allen" (DOMANESCHI, 2016) can mean different things, depending on the intention which the speaker invests it with. Thus, "all of them" can be used to communicate that the speaker does not feel like watching any movie by Woody Allen at all or even that she is a Woody Allen fan, among many other possibilities. Figuring out the communicative intentions of the speaker can only be done if one has enough access to contextual cues.

From the speaker's perspective, in order to communicate she has to produce an effect in the addressee by means of a linguistic expression while making sure that the addressee can identify the intentionality underlying the act if communication. It is interesting to note that, in this perspective, intentionality is linked to the representation of referents. As will be shown below, according to L-AcT, the framework guiding this study, intentionality has more to do with the actional dimension of spoken language, rather than the dimension of mental representations.

Domaneschi (2016) outlines the current trend as to what comprises the different levels of meaning. He proposes that the meaning of an utterance is currently regarded as having four different levels, namely (a) the level of the sentence, (b) the level of what is said, (c) the level of presuppositions, and (d) the level of implicatures. Level (a) encompasses the meaning

conveyed by the linguistic expressions that make up the utterance and the rules that govern their combination; level (b) corresponds to the semantic domain proper, where the linguistic expressions are enriched with contextual information that, for example, disambiguates and saturates the meaning of deictic expressions, determining the truth conditions according to which the proposition conveyed should be assessed. Level (c) is the domain of the information taken for granted, i.e. the information that constitutes the background for what is uttered. Finally, level (d) is the domain of implicatures, i.e. the propositions that are communicated implicitly.

The overview provided above indicates how varied and complex IS in fact is. We will turn now to a review of specific IS categories and terminology. At the end of the chapter, we will try to provide a critical synthesis and briefly introduce L-AcT.

1.3 Information structure: terms and categories

As suggested earlier, terminological consistency across IS frameworks shouldn't be expected. Mathesius (1975[1961]), for instance, calls *theme* the part of the sentence that conveys old information. Halliday's (1967b) *theme*, on the other hand, is defined as the starting point of "the clause as a message", and considerations regarding cognitive status are associated with a different level in the author's systemic approach. While some scholars regard the term topic as the Anglo-American equivalent of theme in the Prague tradition – e.g. Reinhart (1981, p. 3) – others use the terms *theme* and *topic* to refer to different notions; for instance, Halliday (1967, p. 200) says that *topic* inappropriately fuses theme and given.

The following subsections will deal with some of the main IS categories in an attempt to show how varied and even incongruous they may be depending on the approach considered.

1.3.1 Information status

Information status is a cognitive notion that is sometimes established in relation to CG and associated with different linguistic levels. For example, the choice for an expression to designate a referent is often accounted for in relation to what the speaker considers to be present and active in the mind of the addressee at the time of the utterance. So, a referent assessed as unpredictable or new is likely to be designated by linguistic expressions that are very different from those used to designate a referent taken to be known by the hearer, in which case the

referent may not even receive a linguistic form at all. Likewise, unpredicted discourse referents are prone to receiving prosodic prominence, whereas predictable ones, under normal circumstances, are likely to be played down.

Information status underlies other IS notions – e.g. topic/comment, focus/presupposition –, and it is sometimes regarded as a binary or a gradient property. The notion of givenness, for example, applies to discourse entities and describes those entities in terms of new or old information. Scholars use different terminology, sometimes distinguishing old and given, sometimes not. For some, *old* means either that a discourse entity has been mentioned or that it is available in the context. For others, however, *old* is reserved for previously mentioned entities, while *given* means inferable or contextually available – see Arnold et al. (2013).

As already said, information status is associated with sentence position, semantic load of linguistic expressions and prosodic features. But since these formal correlates are also related to other functions, there are some criteria that can be used to make informed guesses as to the information status of discourse referents. These criteria are not always overtly stated in the literature but can be inferred from the works of the scholars that have been referred to so far. The criteria are listed below in order to provide a general picture of how this aspect may be approached. The answers to the questions below are indicated only in terms of old/new from the perspective of the speaker, but other perspectives may apply; the uncertainty of the answers suggests how difficult it is to deal cognitive notions like given and new, showing that the same status can be accounted for, and manifested through, different elements.

1. Is the referent conveyed through a reduced linguistic form? yes (= old, unless the speaker is wrong regarding the CG), no (= probably new)
2. Is it prosodically salient (e.g. has pitch accent, longer duration, or higher intensity)? yes (= probably new), no (= probably old)
3. Is it positioned in the beginning of the sentence or presented in focalized constructions (e.g. cleft or pseudo-cleft sentences)? yes (= probably old), no (probably new)
4. Has the referent been mentioned in the previous discourse or is it available in the communicative situation? yes (= probably old), no (= probably new)
5. While being unmentioned and not present in the context, can the referent be accommodated through some sort of inferential process? yes (= old), no (= new).

Prince (1992) makes an interesting distinction according to which new/given in the discourse model is not the same as new/given to the hearer. In the former case, either prosody alone or a combination of prosody and syntax – e.g. clefts/pseudo-clefts or topicalization – would mark what is old and what is new, like in the example below, adapted from Prince (1992, p. 300-1).

(1.2)

a. It's JOHN I like.

a'. I like x.

a''. x = John.

The sentence in (1.2a) is an example of a focus-presupposition construction, which can be interpreted as the open proposition in (1.2a') plus the instantiation of the variable x in (1.2a'') constituting the focus. An open proposition, following Chomsky (1971), is usually regarded as given or at least somehow present in the discourse model, while the focus is taken to be new.

As for the other type of information status, Prince (1992) identifies it with what the speaker presumes to be present to the addressee. This sort of status is manifested in somewhat simpler ways, e.g. the definite or indefinite form given to the noun phrases in italics below (PRINCE, 1992, p. 303):

(1.3)

a. In the park yesterday, *a kid* threw up on me. [new to the hearer]

b. In the park yesterday, *the kid* threw up on me. [old to the hearer]

Similarly, Erteschik-Shir (2007) draws a distinction between old and given: while old means previously mentioned, given refers to what is present in the mind of the hearer.

In sum, information status is associated with many formal signals, e.g. certain affixes, word order, argument structure choices, referential form, definiteness, prosodic features etc. However, establishing a grammar for these formal signals seems hardly feasible: each element is associated with many different roles and informational functions are established on the basis of criteria that are hard to be consistently applied.

Directly or indirectly, the following papers and books constitute important references to what has been presented in this section: Strawson (1964), Halliday (1967b), Chafe, (1976), Prince (1979, 1992), Vallduvì (1992), Lambrecht (1994).

1.3.2 Topic

The concept of *topic* (HOCKETT, 1958) is closely associated with the notion of theme from the Prague School tradition, just as its counterpart *comment* is closely associated with rheme. As mentioned earlier, some linguists tend to consider topic/comment as North American equivalents for theme/rheme. Usually, topic is related to both the left periphery of the sentence and old information, while the comment is related to the part of the sentence that follows the topic and to new information.

For some scholars, topic is a linguistic expression that establishes a frame for interpretation of the predication (CHAFE, 1976); for others, it is a specific type of theme, i.e. an element that occurs in the beginning of the clause and that refers to a participant, circumstance, or process – e.g. Halliday (1967) and Ping (2004). *Theme*, on the other hand, is associated with elements that play some type of textual function, such as conjunctions and adjuncts do. According to this view, a sentence like “therefore, the city had to be evacuated”, would have “therefore” as theme and “the city” as topic.

For Chafe (1976), the topic of a sentence provides a frame that restricts the domain to which the predication applies. The topic, in this view, must be overtly realized at the left periphery of the sentence, providing a domain – individual, spatial, temporal etc. – to which the predication is anchored.

But Erteschik-Shir (2007, 1997), based on Strawson’s (1964) notion of topic as a pivot for truth-value assessment – i.e. topic as the element according to which the proposition conveyed by the sentence must be judged to be true or false –, argues that even thetic sentences, like those in (1.4) below, have a topic. In sentences of this kind, the here-and-now provides the parameters according to which the truth value of the sentence must be evaluated.

(1.4)

a. It’s snowing.

b. There’s a car outside the door.

The sentences in (1.4), although lacking an overt topic, offer examples of what Erteschik-Shir refers to as *stage topics*, i.e. elements of the 'here-and-now' of the interaction according to which truth-value is assessed (ERTESCHIK-SHIR, 2007, p. 17-8). Stage topics need not be codified linguistically, as they are contextually available. Besides the here-and-now, topics can also be derived from background knowledge provided they are somehow activated, allowing for the accommodation of non-old topics, like the example below shows.

(1.5) John was murdered yesterday. The knife lay nearby.[5]

Erteschik-Shir explains that, if a murder happened, then a weapon was used. Since a knife qualifies as a subset of potential murder weapons, its "topichood status" would be licensed, hence the second sentence in the example above being proposed as acceptable. This illustrates a process whereby unmentioned elements become available eventually being used as topic. This process constitutes a type of accommodation.

As can be seen, topic is a term that is used to refer to concepts that may differ formally and/or functionally depending on the framework considered. But as notions of topic will be further explored in the next two chapters, let us look at other relevant IS notions.

1.3.3 Focus

The notions associated with the term *focus* are frequently established in relation to other IS notions. Bolinger (1954), for instance, identifies the focus of a sentence with the its most informative part, which usually coincides with some type of prosodic prominence. More recently, Vallduvì (1992) and Lambrecht (1994) both point to Halliday (1967b) as a precursor of the way focus is viewed in modern linguistic approaches. This is how Halliday defines it:

> Information focus reflects the speaker's decision as to where the main burden of the message lies. It is one of the many diverse phenomena referred to by speakers of English as "emphasis" (…). Information focus is one kind of emphasis, that whereby the speaker marks out a part (which may be the whole) of a message block as that which he wishes to be interpreted as informative. What is focal is "new" information; not in the sense that it cannot have been previously mentioned, although it is often the case that it has not been, but in the sense that the speaker presents it as not being recoverable from the preceding discourse (…). [Focus] is that which is represented by the speaker as being new, textually (and situationally) non-derivable information (HALLIDAY, 1967b, p. 204).

[5] Roberts (2003, p. 300 apud ERTESCHIK-SHIR, 2007, p. 19).

Lambrecht (1994) presents a somewhat more nuanced view of what constitutes the focus of a sentence. In line with Halliday and Bolinger, his notion of focus is related to new information, but he understands it as a pragmatic relation which cannot be identified with a particular portion of the sentence in a direct way.

(1.6)

a. Where did you go last night?

b. I went to the movies.

Discussing the example above, Lambrecht (1994, p. 209) argues that it would be inaccurate to identify "the movies" as the focus of the sentence (1.6b). This is because, in his account of IS, it is not the referent itself that actually correspond to new information, but rather its role in a pragmatically presupposed proposition – in this case, "the speaker went to x". Thus, "the movie" can only be said to be the focus of (1.6b) if it can be construed as a piece of new information in a pragmatic relation equivalent to the one expressed in "the place I went to last night was the movies" (LAMBRECHT, 1994, p. 210).

More recently, Krifka (2008) and Krifka and Musan (2012) define focus as a function that expresses the presence of alternatives by means of sentence accent, an idea that constitutes a development of Chafe's notion of contrastiveness (CHAFE, 1976) and ideas from the framework of *Alternative Semantics* (ROOTH, 2016). Krifka notes that there are two types of focus: one that can be thought of as simply an information packaging phenomenon (pragmatic focus), and another which affects truth-conditional aspects of discourse (semantic focus). For instance, in the sentence "John [MARried]$_F$ Sue"[6], the focus on "married" indicates alternatives to "marrying", e.g. 'meeting', 'dating', 'getting engaged' etc., having no influence on truth-condition assessment. But with focus-sensitive particles such as *only*, *even*, and *also* the placement of focus has truth-conditional effects. A sentence like "John only introduced Mary to Sue" will have different truth conditions depending on whether it is uttered with focus on "introduced", "Marry", or "Sue".

The notion focus enters the generative grammar with Chomsky (1971) and Jackendoff (1972), who incorporate the pair focus-presupposition to deal with the problem of semantic representation and its relationship with syntax (see ERTESCHIK-SHIR, 2007, MOLNÁR;

[6] Krifka and Musan (2012, p. 7).

WINKLER, 2005). As one would expect from a formal approach of this kind, they seek to embed IS in a theory of language made up of hierarchically organized components.

Jackendoff (1972) incorporates focus as one of the four parts of his theory of semantic representation. These four parts are:

1. the functional structure, where verbally induced relations such as agency, motion and direction take place
2. the modal structure, which "specifies the conditions under which a sentence purports to correspond to situations in the real world" (JACKENDOFF, 1972, p.3)
3. the table of coreference, where pairs of NPs are specified as coreferential or not, and finally
4. the part associated with presupposition and focus, which have to do with the organization of old and new information in the sentence.

For Jackendoff, focus is "the information in the sentence that is assumed by the speaker not to be shared by him and the hearer" (JACKENDOFF, 1972, p. 230). Presupposition, on the other hand, is "the information in the sentence that is assumed by the speaker to be shared by him and the hearer" (JACKENDOFF, 1972, p. 230). The author, in line with Chomsky (1971), advance an account of semantic interpretation that is not entirely determined by deep structure. For instance, in (1.7) below,

(1.7)

a. Is it JOHN who writes poetry?
b. No, it is BILL who writes poetry.
c. No, it is JOHN who writes short stories.

(JACKENDOFF, 1972, p. 229)

what would make (1.7b) an appropriate response to (1.7a), Jackendoff explains, is the fact that both sentences share the same presupposition, i.e. that there exists someone who writes poetry, while (1.7c) presupposes something slightly different, i.e. that there exists someone who writes short stories. The focus of (2.7b) is assigned to Bill, which is realized with particular stress and intonation patterns conveying the newness of the sentence.

Jackendoff (1972) has it that focus assignment is a surface phenomenon implemented by stress, as opposed to the Katz-Postal (1964) hypothesis, which assumes that focus is assigned to the predicate of the higher clause in the deep structure. Jackendoff characterizes focus assignment as follows: "[i]f a phrase P is chosen as the focus of a sentence S, the highest stress in S will be on the syllable of P that is assigned highest stress by the regular stress rule" (JACKENDOFF, 1972, p. 237).

(1.8)[7] Surface structure

a. Did Max kill the judge with a GUN?

b. Did John leave at six o'CLOCK?

(1.9) Deep structure

a. the thing [MAX KILLED THE JUDGE WITH WH-SOMETHING] was a gun

b. the time [JOHN LEFT AT SOME TIME] was six o'clock

A basic problem of a deep structure account of focus assignment, according to Erteschik-Shir (2007), becomes apparent with the sentences in (1.8) and corresponding deep structures in (1.9), where two-clause deep structures have to be posited in order to derive surface structures that are actually made up of one clause alone. Reviewing more recent formal theories, Erteschik-Shir (2007, p. 4) points out that IS has a marginal role in theorizations associated with *Principles and Parameters* (see CHOMSKY, 1986) and that it is hard to establish a coherent account of IS and its relationship with syntax, semantics and intonation within hierarchical frameworks. According to her, theories that allow parallel computations of different linguistic levels are more efficient in mapping IS and the different levels of linguistic organization.

1.4 Information structure according to the Language into Act Theory

The *Language into Act Theory*[8] (L-AcT) regards IS in terms of pragmatic functions inextricably related to prosody (CRESTI; MONEGLIA, 2010). A pragmatic approach to the

[7] The examples in (1.8) and (1.9) are adapted from Jackendoff (1972, p. 231)
[8] Cresti (2000), Moneglia and Raso (2014), Cavalcante (2015).

study of speech, it proposes that no account of spoken language is complete if "actionality", i.e. the illocutive dimension, is overlooked.

L-AcT assumes that there is an inventory of prosodic forms associated with stable informational functions which speakers have at their disposal in order to encode their messages. Information and prosodic units are considered two dimensions of the same object, and classes of prosodic units are systematically associated with functions and positions inside the utterance, which is defined in terms of pragmatic and prosodic autonomy, as will be expounded in detail in the following chapter.

Within L-AcT, terms like comment, topic, and parenthetic refer to concepts that are, in many ways, different from what they refer to within other frameworks. The difference is derived in part from the particular importance of prosody in the theory as well as the understanding that the speech act (AUSTIN, 1962) constitutes one of the most important aspects of spoken language (see chapter 2).

L-AcT recognizes that prosody marks the boundary of the utterance and may segment utterance into mutually dependent parts, giving rise to complex autonomous units made up of components associated with informational functions other than the performance of the illocution. Out of the multiple formal elements that constitute an utterance, prosody also stands out for its importance as an information marker, thus signaling the pragmatic autonomy and communicative goal of the utterance.

Information structure, according to L-AcT, has to do with the construction of utterances in terms of information units defined with respect to prosodic characteristics, position in the utterance, and informational function. The definition of each unit is independent of participants' knowledge or attentional focus, recency of mention of entities or aspects of the like. As a matter of fact, information units are in a sense independent of their locutive makeup, although certain units tend to correlate with certain types of syntactic/semantic content.

Looking at a spoken text from the perspective of L-AcT is equivalent to trying to grasp how prosodic units are mapped onto functions that are centered around the illocutionary act. These units can be textual, when their role has to do with the syntactic/semantic makeup of the utterance, or dialogic, when they have a regulatory role in the interaction and address the interlocutor. Within other frameworks, dialogic units correspond to discourse markers, and the IS model developed within L-AcT is able to account for discourse markers as well. Each of those units, both the textual and dialogic ones, is marked prosodically and distributionally.

Let us consider, for example, the information unit of Topic (TOP), which is the focus of this study. Its function is to supply a cognitive domain according to which the illocution – not the predication or the proposition, as in other frameworks –, carried by the information unit of Comment[9] (COM), must be interpreted. TOP is a category that applies only to speech, as it is indissociable from certain prosodic parameters.

According to Cresti (2011, p. 49), other frameworks tend to rely too much on the context and thus fail to distinguish between *the pragmatic level* of the illocution – a product of an unconscious drive – and *the linguistic level* – a product of cognitive computations. The context, Cresti (2011) argues, important as it is for communication, does not determine language, a system that can be used largely independently of the context. In this view, there is an unbridgeable gap separating contextual stimuli from the affective response that constitute the illocution present in every utterance. Accordingly, the illocution – the most informative aspect of the utterance – cannot be thought of as a mere predictable reaction. Therefore, to regard the concept of topic in terms of old information expressed by a contextual element linguistically codified in an utterance, or as an instruction for the entering of information into the CG, requires that language be considered, to a greater or lesser extent, inseparable from context. However, empirical evidence shown by studies carried out within L-AcT (see CRESTI, 2011) shows that one of the most relevant functions of TOP is precisely to allow for context displacement – be it spatial, temporal, or existential –, an aspect considered by Hockett (1960) as one of the defining features of natural languages.

Thus, within L-AcT, TOP is said to codify a cognitive rather than a contextual domain. According to Cresti (2011), the context is always an open set that cannot be properly represented in terms of a closed set of a logical universe.

In this view, IS constitutes a level that, in a sense, pairs morphosyntactic makeup with communicative value, thus giving rise to an utterance. To give the simplest of examples, a proper name – say, "John" – can only be interpreted as an act of communication – or part thereof – if the prosodic features with which it is uttered is taken into account. Let us consider the utterances in the example below.

(1.10)

---

[9] "TOP" and "COM" are the abbreviations used within L-AcT to refer to topic and comment. In this dissertation, they refer exclusively to how these terms are defined within L-AcT.

a. John is an awesome guy //=COM= %assertion[10]

b. John /=TOP= he is an awesome guy //=COM= %assertion

c. he is an awesome guy /=COM= John //=ALL= %assertion

d. John //=COM= come on over here now //=COM= %calling and %order

The utterance in (1.10a) is organized as a single prosodic unit carrying a sentence in which "John" is the subject. In utterance (1.10b), "John" is the locutive makeup of a topic unit, thus constituting an individual domain in relation to which the remainder of the utterance, the illocution, must be interpreted. In order for it to be a TOP unit, it must be realized in a dedicated prosodic unit characterized by specific patterns of melodic variation and syllable lengthening (see chapters 4, 5, and 6). As a subject, "John" is an argument of the sentence in (1.10a); as a TOP, "John" supplies a cognitive domain for the interpretation of the illocution in (1.10b), thus establishing a pragmatic relation of aboutness to the COM unit.

In the utterance in (1.10c), on the other hand, "John" is the locutive content of a dialogic unit, more specifically, an allocutive (ALL). The function of this unit is to identify the addressee, in addition to possibly promoting social cohesion (see chapter 2). As a dialogic unit, it depends on the utterance which it is a part of, as it cannot receive a pragmatic interpretation if considered in isolation.

Finally, in (1.10d) "John" constitutes the entire locutive content of an utterance conveying an act of calling. In this example, "John" has pragmatic autonomy. The interpretation of a mere proper noun, as seen in these examples, is subjected to the interpretation of the information unit that carries it. In fact, from the point of view of communication, little can be said about any piece of uttered locutive content without recourse to its prosodic characteristics.

It must be acknowledged that other approaches have contributed immensely to advancing the understanding of the interplay between language and pragmatics. As Arnold et al. (2013) note, scholars working within a variety of frameworks generally agree that language form is sensitive to IS. Formalists at least since Chomsky (1971) have incorporated IS notions such as focus and presupposition to account for linguistic phenomena that cannot be explained by syntactic considerations alone. More recently, psycholinguists have shed light on IS and how psychological mechanisms are involved in the production and interpretation of language – see Arnold et al. (2013) for a brief review).

---

[10] The sign "%" precedes the specification of illocutionary value.

The aim of this chapter was to review some of the major contributions by different approaches to the study of IS, putting L-AcT into perspective and pointing to the pertinence and empirical validity of the framework, for which evidence will be provided in the next chapter.

Rather than reject or criticize IS interpretations other than that by L-AcT, the multiplicity of concepts is regarded here as a result of the complexity of the subject and as an indication that this area of inquire is still rich and lively.

Finally, it must be said that the L-AcT model of IS provides a scheme for the annotation of informational functions, which makes it possible for speech to be studied in a coherent way. These functions are covered in the next chapter. In chapter 5 we will show how reliably can these functions be detected by discussing the results of an agreement test using the kappa statistic (COHEN, 1960, FLEISS, 1971).

2 The Language into Act Theory

The Language into Act Theory (L-AcT) is the theoretical framework that guided this study. L-AcT was first formalized in Cresti (2000), following decades of extensive research on spontaneous speech data collected in corpora. It was developed at the *Research unit of the Humanities Department of the University of Florence* (LABLITA)[11], led by E. Cresti and M. Moneglia.

L-AcT constitutes a pragmatic approach for the study spontaneous speech that stems from *Speech Act Theory* (AUSTIN, 1962) and that draws in part on the prosodic model developed at the *Institute for Perception Research* (IPO), Eindhoven, presented in 't Hart et al. (1990). It is a theory grounded on empirical data collected in carefully designed corpora

Numerous studies have been carried out within the L-AcT framework, many of which at the *Laboratory for Empirical and Experimental Linguistic Studies*[12] (LEEL), where researchers have been actively contributing to the theoretical and methodological advancement of L-AcT since the year 2008. This contribution has been made both in the form of studies that look at language according to the L-AcT tenets as well in the form of compilation and treatment of speech corpora. With respect to the latter, the C-ORAL-BRASIL corpora I and II (RASO; MELLO, 2012, RASO; MELLO; FERRARI, forthcoming) and the AE minicorpus (see chapter 4 and CAVALCANTE; RAMOS, 2016) are particularly relevant for the present work. Currently, a corpus of speech produced by schizophrenic patients, the C-ORAL-ESQ (RASO; MELLO; FERRARI; ROCHA, in preparation), is being compiled according to the L-AcT methodology.

This chapter deals with L-AcT, presenting the underlying assumptions and theoretical constructs of this framework.

2.1 The reference unit for speech analysis

L-AcT is based on the notion that pragmatic autonomy in speech is fundamentally marked by prosody. Spoken discourse is understood as a series of autonomous units called utterances consisting of at least one prosodic unit each. The prosodic units may have different functions,

[11] LABLITA website: http://lablita.dit.unifi.it/.
[12] LEEL website: http:// http://www.letras.ufmg.br/leel/.

depending on their formal characteristics, but every utterance contains at least one mandatory unit that conveys the speech act (AUSTIN, 1962), a core concept in the theory.

Regarded in terms of speech act performance, the utterance consists of three simultaneous, parallel acts: the locutive, the illocutive, and the perlocutive acts. The *locutive act* corresponds to the morphosyntactic and semantic levels of the utterance. The *illocutive act* corresponds the speaker's ultimate communicative goal in producing an utterance; its formal signal pertains to the prosodic level. Illocutive acts are translatable as assertions, questions, greetings and other various acts (see MONEGLIA, 2011 for a tentative typology of illocutions). Finally, the *perlocutive act* is described as the libido asset that drives the emergence of the speech act as a whole[13].

As mentioned earlier, the utterance is organized around a prosodic unit whose pragmatic function is the performance of the speech act. This unit can be surrounded by other units playing different pragmatic roles related to either the textual makeup of the utterance or the regulatory dimension of the interaction. In the former case, they are called *textual units*; in the latter, *dialogic units*. These two macro categories are picked up in the next section.

A prosodic unit is delimited from its surrounding neighbors by a discontinuity perceived as an acoustic boundary (RASO et al., 2015, TEIXEIRA, 2018) and referred to as a *prosodic break*. A prosodic break can basically have two functions: (i) to mark the completion of the utterance or (ii) to separate units that are part of the same utterance. The prosodic break that marks the completion of utterances is perceived by the hearer as conclusive and called *terminal prosodic break*. On the other hand, the break that marks the division between prosodic units inside the same utterance is perceived as continuative and called *non-terminal prosodic break*.

The perceptual relevance of prosodic breaks is shown by studies that report very good results achieved in inter-rater agreement tests using the kappa statistic[14] to measure the reliability of the detection of terminal and non-terminal prosodic breaks (MONEGLIA et al., 2010, RASO; MITTMANN, 2009, MELLO et al., 2012). The mean kappa coefficient for the detection of terminal and non-terminal breaks reported in these studies is 0.86, which is considered an excellent result, according to Landis and Koch's (1977) benchmarks (see chapter 5).

---

[13] Note that the definition of perlocutive act by L-AcT differs from that originally proposed by Austin (1962), who describes the perlocutive act as the effect of the speech act upon the speaker's audience.
[14] Cohen (1966), Fleiss (1971)

In the L-AcT annotation scheme, terminal breaks are indicated by a pair of forward slashes ("//") while non-terminal breaks are indicated by one single slash ("/"). The two following examples show the clearly different perceptual effects of these two types of breaks.

(2.1) afamdl02_172 (ex2.1.wav)[15]:
*DAR: you have no idea //

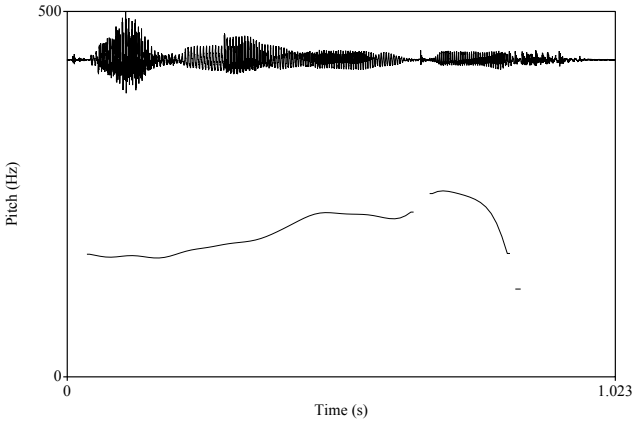


Figure 2.1 – Waveform and *f0* contour of the utterance in ex. (2.1)

By listening to *ex2.1.wav* in the accompanying folder, one can appreciate the pragmatic autonomy of the utterance in (2.1), even if determining its precise illocutionary value is not as straightforward one might wish. Consisting of one prosodic unit alone, this is an instance of a *simple utterance*. The next example shows a different case.

(2.2) afammn01_10 (ex2.2.wav):
*LYN: in the beginning of the year / there was a lot of classroom work //

[15] The examples shown in this and the following chapters come from the C-ORAL family annotated minicorpora (see chapter 4). They are introduced by a code identifying the minicorpus and text from which they were taken followed by the identification of their corresponding audio file in the accompanying folder. The first letter in the code indicates the language of the minicorpus (a = AE, b = BP, p = PE, and i = IT), the following three indicate the domain of interaction (fam = family/private, pub = public), and the final two letters indicate the type of interaction (mn = monologue, dl = dialogue, cv = conversation). The numbers represent the rank of the utterance in the text from which it was taken.

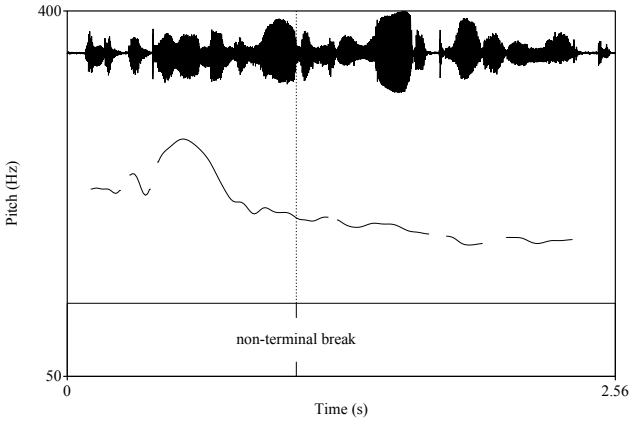


Figure 2.2 – Waveform and *f0* contour of the utterance in ex. (2.2)

The vertical dashed line in Figure 2.2 above indicates the point in time of the prosodic boundary inside the utterance. By listening to *ex2.2.wav*, one can readily notice that this utterance consists of two different, interrelated prosodic units: one hosting the string "in the beginning of the year" and the other hosting "there was a lot of classroom work". An utterance like this, made up of more than one prosodic unit, is called *compound utterance*.

The continuity effect of the non-terminal break in (2.2) becomes particularly clear if the first unit is heard in isolation (*ex2.2a.wav*). Such is not the case with the second unit which, even if heard in isolation (*ex2.2b.wav*), maintains the ability to convey its pragmatic function autonomously. It will be shown below that this autonomy is imparted mostly by prosody, rather than by aspects of syntactic and semantic nature.

2.1.1 Alternative criteria for utterance detection

The definition of utterance based on the notion of pragmatic autonomy signaled by prosody represents an advance attained after years of empirical research (CRESTI, 2000; MONEGLIA, 2011; MONEGLIA; RASO, 2014). Traditional approaches to the study of spoken language have relied heavily – and often exclusively – on categories typical of written-language data (RASO, 2013; MELLO, 2014). As a consequence, attempts to establish reliable criteria for retrieving coherent units from spoken discourse have frequently yielded debatable results.

The notion of utterance as *a spoken sentence*, for instance, has been adopted by frameworks of formal and functional orientations, where *sentence* is defined as either the verbal predication of a noun phrase or the maximum projection of a verbal nucleus. But by operating

with a syntactically oriented definition of utterance, one risks leaving a considerable amount of data unaccounted for, since, on average, 30% of speech utterances actually have no verb at all (CRESTI, 2005, BIBER et. al, 1999, RASO; MITTAMNN, 2012; RASO et al., 2015), not to mention the utterances containing verbs that are not the heads of the highest syntactic phrases. If these utterances are considered together with the verbless ones, over 50% of them fall short of being captured by the notion of utterance as a spoken sentence.

An alternative to that syntactically oriented definition is the notion of utterance as *a dialogic turn*. This, however, brings forth considerable problems as well, given the remarkable heterogeneity that dialogic turns exhibit. A dialogic turn, for instance, may be as short as an interjection lasting a few hundred milliseconds, or as large as a stretch of speech lasting several minutes. Different studies have shown how dialogic turns lack the necessary uniformity to be considered practical minimal units for speech analysis (see, e.g., CRESTI; GRAMIGNI, 2004; MELLO, 2014, RASO et al., 2015).

Another alternative yet would be to take the pause – defined as an acoustic interruption in the speech signal larger than a threshold somewhat arbitrarily defined – as the demarcation of the utterance. But the pause criterion, despite its appealing objective nature, has proven inappropriate as a reliable strategy for utterance detection at least for two reasons. On the one hand, many utterances, like the ones in (2.3) below, have no pause of any kind separating them from a neighboring utterance. That is, having a pause is not a necessary condition for terminal prosodic.

Actually, prosodic breaks and pauses are different phenomena that mustn't be equated or indiscriminately paralleled (see Raso *et al.,* 2015). To be sure, whenever the speaker produces a pause, a prosodic break is also produced, but prosodic breaks neither entail nor require the presence of a pause. In (2.3) below, the discontinuity in the *f0* curve at the boundary separating the two utterances is due to the two contiguous voiceless segments – [f] and [ʃ] – at the end of the first utterance and in the beginning of the second one, respectively.

(2.3) afamdl03_7-8 (ex2.3.wav):
*ALC: so she's / all of 'em / are not short enough // she's / these are the shorter one / and they're about two inches too long hhh //

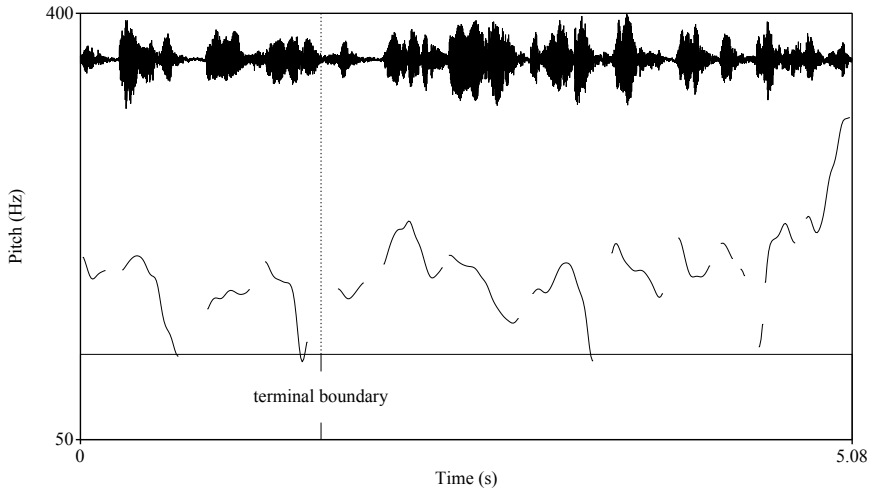


Figure 2.3 – Waveform and *f0* contour of the utterances in (2.3)

The other reason that renders the pause inadequate as a criterion for identifying utterances is that there are many utterances that contain a pause as part of their internal prosodic structure, as (2.4) below shows.

(2.4) afamdl02_99 (ex2.4.wav):
*PAM: this incredible / film legend //

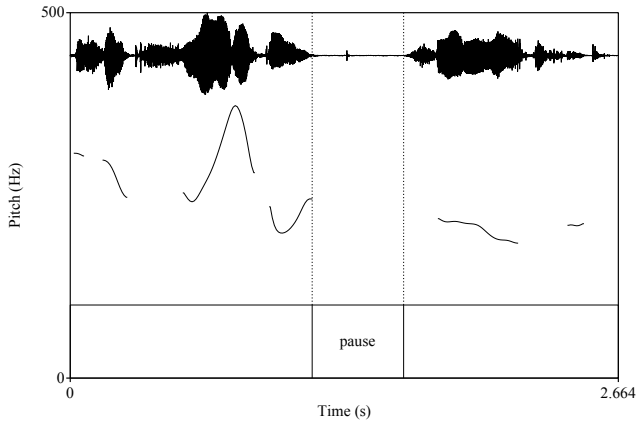


Figure 2.4 – Waveform and *f0* contour of the utterance in (2.4)

In sum, to take the pause as a criterion for identifying utterances creates two types of distortions. On the one hand, adjacent utterances that are not separated from each other by a pause simply fail to be detected and are counted as one. On the other hand, those utterances that contain internal pause(s) end up being counted as two (and possibly more).

Raso *et al.* (2015) studied the possible association between (i) pause and terminal break and (ii) pause and non-terminal break, seeking also to verify the possible association of pause duration and type of prosodic break. Their results strongly suggest that pause cannot be taken as a predictive feature of either terminal or non-terminal breaks, and that it is not possible to establish a consistent association between the duration of the pause and the type of the break.

In light of what has been discussed in this section, the L-AcT proposal that utterances be identified on the basis of a pragmatic autonomy marked by prosody stands out as an interesting and empirically valid alternative. Details regarding speech segmentation in accord with L-AcT can be found in Raso et. al (2015), Mittmann and Barbosa (2016), Teixeira and Mittmann (2018), and Teixeira (2018). The latter study, which examines an array of acoustic-phonetic parameters for the production and perception of prosodic boundaries in spontaneous speech using a complex statistical methodology, has proposed a model for automatic detection of terminal prosodic breaks that shows 80% of convergence with detection by human annotators. Teixeira (2018) also dealt with non-terminal prosodic breaks, but these constitute a more challenging phenomenon for which she proposes three different models, each of which exploring specific combinations of phonetic measurements capable of detecting varying proportions of non-terminal breaks.

In the next section, the L-AcT account of information structure will be discussed.

2.2 The relationship between prosodic and information units

In chapter 1, some of the major approaches and concepts related to IS were outlined, and some of the basic assumptions of L-AcT regarding the subject were presented. In this chapter, we will delve into IS according to this framework.

As mentioned earlier, L-AcT proposes that the units of the prosodic pattern are systematically associated with information functions, the primary of which being the conveyance of the illocution. In this view, classes of prosodic units are apt to fulfil specific pragmatic roles within the host utterance. L-AcT draws on the IPO model for the perceptual study of intonation (t' HART *et al.*, 1990), focusing on acoustic features that are perceptually salient so as to provide a coherent account of what is linguistically relevant.

From the phonetic point of view, the prosodic unit is regarded as a configuration containing at least one pitch movement – i.e. one perceptually relevant *f0* movement. Only one type of configuration is found to be obligatory in every utterance. This configuration is referred

to as *root*. The root unit may be preceded, followed, or even interrupted by other optional prosodic units, and the possibilities for the arrangement of pitch configurations are numerous. The relationship between prosodic units and information units is accounted for by the *Informational Patterning Hypothesis* (CRESTI; MONEGLIA, 2010, CRESTI, 2011), which establishes a parallel between prosodic units and informational roles. In the next section, the information units and their general prosodic features will be presented.

2.3 The units of the information pattern

Information units can be of two macro types, as mentioned earlier, depending on the general nature of their function. There are the so-called *textual information units*, which make up the locutive content proper of the utterance, and the *dialogic information units*, which are involved in the regulation of the verbal exchange, being typically directed at the interlocutor. Dialogic units correspond to what is referred to in other frameworks as discourse markers – for details, see Raso (2014).

Information units are defined on the basis of three criteria: (i) position inside the utterance, (ii) prosodic characteristics, which may involve different parameters, e.g. *f0* movement, syllable lengthening etc. and (iii) pragmatic function (CRESTI, 2000; MONEGLIA, RASO, 2014).

Syntactic and semantic aspects are not considered central for characterizing information units. In fact, Cresti (2000, 2014) regards these units as islands which, apart from certain units dedicated specifically to textual integration – see the case of the appendix of comment and appendix of topic below–, do not maintain strict syntactic-semantic relationship with one another. According to Cresti (2014), "the syntax of the utterance *does not* correspond to a unitary hierarchical configuration, but to the combination of *local* syntactic clauses, phrases, or fragments" (p. 368, emphasis added). This view, however, is sometimes challenged by some empirical evidence which suggests that syntactic compositionality between information units may in fact occur, even if they are not necessary (see SILVA, 2019).

The group of textual information units comprise the comment (COM), the bound comment (COB) and the multiple comment (CMM), the appendix of comment (APC), the topic (TOP), the appendix of topic (APT), the parenthetic (PAR), and the locutive introducer (INT). The dialogic units comprise the incipit (INP), the phatic (PHA), the expressive (EXP), the allocutive (ALL), the conative (CNT), and the discourse connector (DCT).

These units will be described in the following sections. Focus will be given to the units that have textual functions, as they are more closely related to this study. As for the dialogic units, Raso and Moneglia (2014) constitute a good introduction. For a comprehensive discussion, see Raso (2014), Raso and Vieira (2016) and Gobbo (2018).

2.3.1 The comment (COM) and its appendix (APC)

The function of COM is to carry the illocutionary force, thus conveying the primary communicative value of the utterance. The unit has a prosodic prominence called *nucleus*, which communicates the its function. The syllables (or syllable) of the nucleus constitute the *prosodic form* of the unit, an arrangement of prosodic parameters consistently associated with conveying the same type of pragmatic function (FIRENZUOLI, 2003).

A prosodic form[16] need not (and usually does not) correspond to the entire syllabic extension of the unit. Therefore, a prosodic form is frequently accompanied by a *preparation*, i.e. a portion that precedes the nucleus and that has no information function; a *coda*, like the preparation but occurring after the nucleus, and sometimes even a *linking portion*, joining two discontinuous parts of the same nucleus. Linking portions have been widely observed in TOPs, which are discussed below.

The prosodic characteristics of COM vary depending on the type of illocution carried by COM. It is subject to pragmatic and cognitive parameters, which are particularly important in cases where the prosodic characteristics of the unit are not sufficient to determine its pragmatic interpretation. In addition, it is likely that the number of illocutions exceed the number of prosodic forms. So, two illocutions that are carried by the same prosodic form are disambiguated by cognitive and pragmatic parameters. The precise identification of parameters that are relevant for the interpretation of each type of illocution is an endeavor which has been experimentally pursued in Rocha (2016).

Empirical studies based on spontaneous speech corpora have pointed to different types of illocutions (MONEGLIA, 2011). Although the precise number of illocutions has not been determined thus far, it certainly exceeds the number of illocutions proposed in the context of logic-based frameworks, such as the one found in Searle (1975).

---

[16] The topic (TOP) is the other information unit that has a prosodic form.

COM is a unit with few distributional constraints. It can be the only information unit in an utterance and can be preceded or followed by almost any other information unit, with APC representing an exception. In some particular cases, COM can even be interrupted by another information unit, most frequently a parenthetic (see below).

As for APC, its function is to integrate the textual content of a COM unit, supplying additional syntactic-semantic material that often correspond to given information. It is realized by a prosodic unit of the suffix type and its *f0* contour may show either a flat or a falling *f0* profile. Unlike COM, APC does not have a prosodic nucleus, so its omission does not compromise the pragmatic interpretability of the utterance. As for its position, it always follows the COM unit.

(2.5) afamcv04_10 (ex2.5.wav, ex2.5a.wav, ex2.5b.wav):
*MAR: only the Sam's Club cans /=COM= get so warped //=APC=

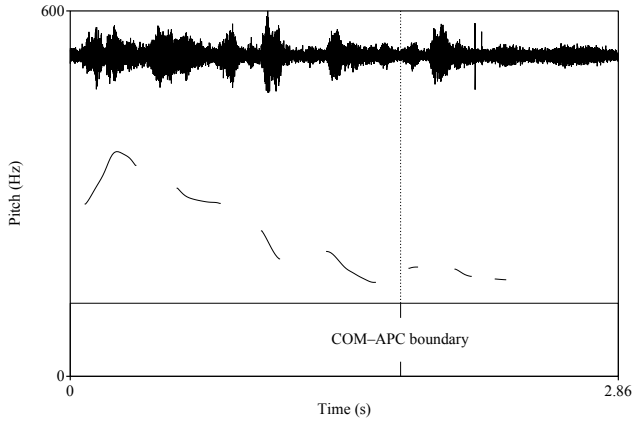


Figure 2.5 – Waveform and *f0* contour of the utterance in ex. (2.5)

Example (2.5) shows an utterance exhibiting a COM–APC information structure. By listening to *ex2.5a.wav* and *ex2.5b.wav*, one can appreciate that the illocutionary force and pragmatic autonomy of the utterance is conveyed by the first unit, i.e. the COM unit. The APC unit, on the other hand, serves uniquely as a textual integration for the content in COM, and therefore is not pragmatic autonomous.

APC supplies syntactic-semantic material to the locutive content of COM, thus aiding in the interpretation of COM, whose locutive content may otherwise be deemed semantically incomplete. The APC does not contribute to the realization of the illocutionary act, and the functional ability of COM remains intact if the APC is omitted.

2.3.2 The topic (TOP) and its appendix (APT)

The TOP unit has the function of supplying a domain of identification – spatial, temporal, individual etc. – for the interpretation of the illocutionary force carried by COM. Borrowing the analogy that Chafe (1976) uses to explain his notion of topic[17], the definition of TOP proposed by L-AcT can be thought of as the "hitching post" for the illocution. An illocution whose application is not being restricted by the presence of a TOP unit must be interpreted according to a contextually given domain.

TOP allows an utterance to be distanced from its immediate context, and a domain present in the context can too be codified as TOP. This may occur for disambiguation purposes, for instance, when an illocution is potentially interpretable according to more than one contextually given domain.

Concerning the position of the unit, TOP always precedes the illocutionary unit to which it serves as a hitching post, with the possible interposition of other units, which include but are not limited to APT.

Despite its lacking the pragmatic autonomy of COM, TOP exhibits a functional prosodic nucleus consisting of a few of its syllables, which are dedicated to imparting its pragmatic function. Earlier studies identified four different types of nuclei – also referred to as *prosodic forms* – in TOP units (FIRENZUOLI; SIGNORINI, 2003, SIGNORINI, 2005, MITTMANN, 2012, ROCHA, 2012). More recently, though, Cavalcante (2015) and Raso et al. (2017) have shown that two of the prosodic forms previously identified actually comprise variations of the very same form.

These studies have not been able to find functional differences among the prosodic forms. They have not been able to establish an association between prosodic form and variables like syllabic extension and semantic content of TOP either.

The prosodic forms are described as follows:

- Type 1:

  Nucleus position: last stressed syllable – and post-stressed ones, if any;

  f0 movement: rising-falling;

---

[17] Chafe (1976) defines *topic* as a frame that restricts the domain within which the predication holds (see chapter 3). The abbreviation TOP is reserved for the information unit of topic according to L-AcT.

Presence of syllable lengthening.

– Type 2:

Nucleus position: last stressed syllable – and post-stressed ones, if any;

f0 movement: rising;

Presence of syllable lengthening.

– Type 3:

Presence of two semi-nuclei, often discontinuous;

First semi-nucleus features high to extra-high *f0* values;

Second semi-nucleus features lower *f0* values;

The two semi-nuclei may be separated from each other by functionally inactive syllables called *linking portions*;

Presence of syllable lengthening.

In AE, some TOP units have been found which, despite their exhibiting no relevant *f0* movement, do cause the perception of pitch variation. Referred to as *Flat TOPs* in Cavalcante (2015), this particular realization of TOP has prompted discussions about the possible existence of a fourth class of prosodic form. The fact that Flat TOPs exhibit two semi-nuclei instead of an indivisible nucleus like Types 1 and 2 do, the possibility that Flat be a subgroup within Type 3 has also been considered. This discussion will be picked up later in chapters 5 and 6, where more will be said not only about Flat TOPs but also about the other prosodic forms of TOP.

Turning now to a particular semantic facet of TOP, some units exhibiting modal function have been found in the C-ORAL family corpora (MITTMANN, 2012, MONEGLIA; RASO, 2014, CAVALCANTE, 2015)[18]. The locutive material of such TOPs expresses the speaker's attitude towards the propositional content expressed in COM, as the example below shows.

(2.6) apubmn01_317 (ex2.6.wav):
and /=AUX= *I think* /=TOP= you know / for &s / *for some of us* /=TOP= underwater viewing gallery has not been / your favorite place /=COB=[19]

[18] For modality and topic in other frameworks, see Chafe (1994) and Mithun (2016).
[19] The tags of the annotation scheme used within L-AcT is fully presented in chapter 4.

The TOP units in (2.6) convey epistemic meaning, expressing the degree of confidence of the speaker as to the propositional content in COB, a particular type of illocutionary unit called *Bound Comment* that features a prosodic signal of continuity and that is presented in section 2.3.6 below.

Modal TOPs may be regarded as a theoretical problem for L-AcT. This is because a modal TOP may be considered to fail to supply an identification domain for the interpretation of the illocution, given that it expresses the speaker's attitude not towards the illocution but rather the proposition that the illocutionary unit conveys. In that sense, modal TOPs would infringe on the principle of non-compositionality between information units, which are traditionally conceived as non-compositional syntactic and semantic islands (CRESTI, 2014).

Nevertheless, modal TOPs are prosodically realized just like ordinary TOPs are, an aspect that has secured their place inside the same group of information units. Furthermore, it has been suggested that modal TOPs do supply a domain of identification, which would correspond to the speaker herself. The problem, however, is still open and lies beyond the scope of this study.

As for APT, it has basically the same function as APC, i.e. it adds syntactic-semantic content to the text in the TOP unit. The prosodic characteristics of APT, however, seem to be a little more complex than those of APC, given that it sometimes reproduces the *f0* movement of the TOP unit, albeit with a range smaller than that of the TOP unit and without the nucleus. APC is always found following a TOP unit, as (2.7) below shows.

(2.7) afamdl02_53 (ex2.7.wav and ex2.7a.wav):
*PAM: the things I know most /=TOP= about life and death /=APT= come from [/1]=SCA= from /=SCA= my grandmother //=COM=

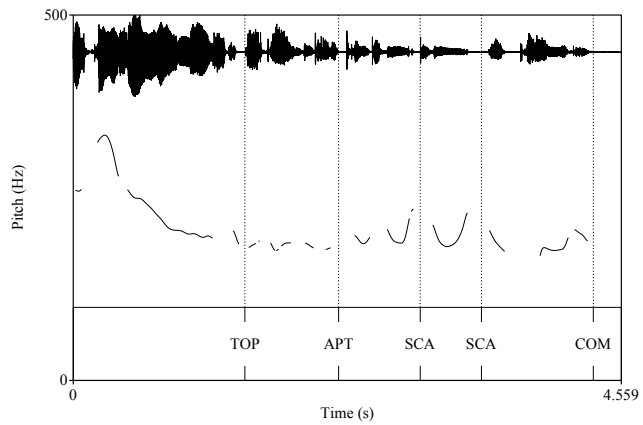


Figure 2.6 – Waveform and *f0* contour of the utterance in ex. (2.7)

The example above shows an utterance with a sequence of three information units, namely TOP, APT, and COM. The units identified as SCA (scanning unit) constitute part of the COM. The role of SCA is to carry part of the locutive content of textual units which the speaker, for some reason, could not – or chose not to – fit into one single prosodic unit (for details about SCA, see section 2.3.7 below). If the APT and SCA units are omitted (as in the *ex2.7a.wav* audio file), both the TOP and the COM units still keep their interpretability, and the utterance holds its pragmatic interpretability.

Turning back to TOP itself, the unit also occurs in compound patterns other than the TOP-APT pattern just shown. Chains of up to four TOP units have been found in IT, PB, and AE. Moreover, there is still a more complex phenomenon, which involves usually two, or occasionally three, TOP units that together constitute one single domain for the interpretation of the illocutionary force. These constitute the so-called *List of topics* (TPL), which may exhibit discontinuous nuclei realized in two separate prosodic units (see FIRENZUOLI; SIGNORINI, 2003, MITTMANN, 2012, ROCHA, 2012, CAVALCANTE, 2015).

(2.8) afamdl05_68 (ex2.8.wav):
*JEN: so /=AUX= if I throw an ace of diamonds /=TPL= and somebody doesn't have any diamonds /=TPL= they're gonna throw a heart on me /=COB= and I'm gonna have to take that <trick> //=COM=

The example above shows a List of Topics made up of two TPLs which together supplies the domain for the illocutionary force carried by the COB unit to interpreted. Note that the illocution in the COB unit applies to a circumstance which is determined by both TPLs, and that considering any of them without the other is insufficient to construe the domain to which the illocution applies.

2.3.3 The parenthetic (PAR)

PAR is an information unit that has a metalinguistic role, as it provides supplemental information to offset imprecisions or lack of information that can potentially render the interpretation of the utterance difficult (TUCCI, 2004, 2010, MENDES, in preparation). PAR often has a modal index in the form of adverbials and clauses conveying epistemic and deontic

meaning. Regarding its distribution, the only position where the PAR does not occur is the absolute beginning of the utterance. There is a functional similarity between modal PARs and modal TOPs, but the prosodic and distributional differences that they show prevent their being considered the same type of information unit.

Typically realized with an overall flat *f0* profile, PAR usually does not exhibit relevant *f0* movements. But when it is realized with particularly low *f0* values relative to its surroundings, it usually exhibits a rising *f0* movement at the end. Also, it may exhibit prosodic transitions in its initial and final portions. The transition in the beginning serves the function of taking the unit to an *f0* level other than that of the utterance, while the transition at the end shifts the *f0* back towards its original level.

Another interesting characteristic of PAR is that it may temporarily interrupt another textual unit, which is eventually taken up once the PAR has been completed. The example below illustrates this phenomenon.

(2.9) afamdl04_173 (ex2.9.wav and ex2.9a.wav)
*JUL: we have several scars /=i-COM= in this barn /=PAR= from people bouncing off of trees hhh //=COM=

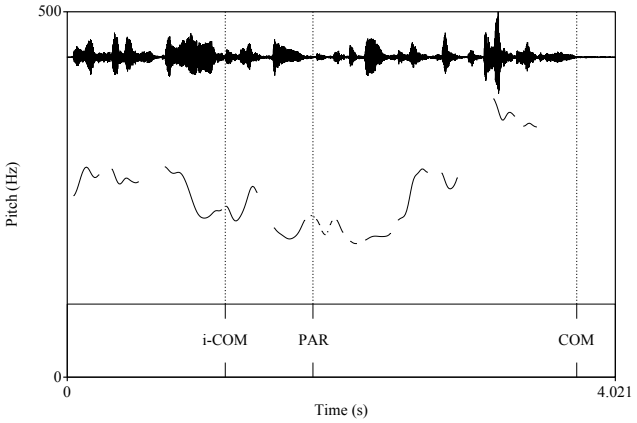


Figure 2.7 – Waveform and *f0* contour of the utterance in ex. (2.9)

In (2.9) above, the label "i-COM" indicates that the COM unit has been momentarily suspended. The that i-COM and COM are parts of the same information unit can be appreciated by listening to *ex2.9a.wav*, from which the PAR was removed. This phenomenon has been observed not only with COM but also with other information units, but only those of the textual type.

Finally, sometimes PAR appears to be strong enough as to suggest that some sort of illocutionary function is being conveyed. Behaving as "an illocution about an illocution", this type of unit has been informally referred to as *peri-illocutionary* PAR. An example is given in (2.10).

(2.10) afammn02_85 (ex2.10.wav):
you know /=AUX= okay /=AUX= your shoe's like this /=CMM= you stretch this out /=CMM= well then it's gonna make these go way out too /=COB= and they get like this /=PAR= then you have to round these back /=COB= you know //=COM=

The illocutionary strength of this type of PAR becomes particularly noticeable when the unit is heard in isolation (*ex2.10a.wav*). No study thus far has looked at PARs taking this particular aspect into consideration. This aspect appears to suggest that an utterance may exhibit, together with the illocutionary and meta-illocutionary level (see the following subsection on INT), a third discourse level. But this is a topic for future research.

2.3.4 The locutive introducer (INT)

INT has the function of signaling that at least one of the information units that come after it must be interpreted according to pragmatic coordinates other than those of the unfolding utterance (MAIA-ROCHA; RASO, 2011). Thus, INT announces a leap to a different hierarchical level.

The unit frequently introduces meta-illocutions such as reported speech, spoken thought, emblematic exemplification, lists, etc., but it may simply introduce content that the speaker chooses to highlight for some particular communicative purpose.

INT is characterized by a falling *f0* movement and an articulation rate that tends to be higher than that of its surroundings. Sometimes the units following an INT unit show an *f0* adjustment leading to a higher melodic level, which is considered to be a sort of index of the change of hierarchical level. As for distribution, INT always precedes the information pattern that it introduces.

(2.11) afamcv03_124 (audio ex11.wav):

*TOC: <I said> to my father /=COB= I said /=INT= hey look /=COB_r=

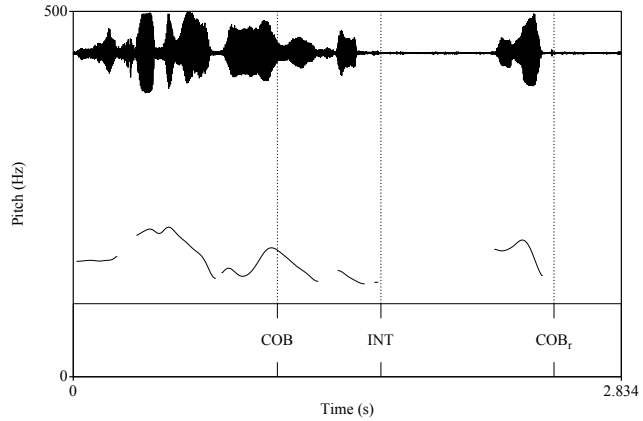


Figure 2.8 – Waveform and *f0* contour of the utterance extract in ex. (2.11)

The INT unit in (2.11) introduces a piece of reported speech – hence the "_r" label accompanying the COB label. The COB unit, as already mentioned, is a particular type of illocutionary unit (see section 2.3.6 below). Being a meta-illocution, the COB unit in the example has to be interpreted according to spatial, temporal and personal parameters different from those of the enunciation itself.

2.3.5 The multiple comment (CMM)

Certain utterances have more than one illocutionary unit. This section presents the illocutionary pattern composed of CMMs.

CMM is a unit that, together with at least another CMM – and possibly other units – constitute what is known as *compositional illocutionary pattern* (PANUNZI; GREGORI, 2012, MONEGLIA; RASO, 2014). A chain of CMMs constitutes a rhetorical arrangement marked by a conventionalized melodic pattern. The melodic characteristics and the rhetorical effect associated with chains of CMMs suggest that these illocutionary units may be the result of a single behavioral program, possibly stemming from a common subconscious drive.

The rhetorical effects created by CMMs can be of different kinds, constituting an open set of possibilities. These effects may join illocutions with one another, express conventionally patterned lists and comparisons, coordinate and modulate requests, strengthen assertions, and even provide a melodic package for alternative or double directives (PANUNZI; GREGORI, 2012, p. 137-8). The patterns in the utterances below illustrate some of these effects.

(2.12)

a. afammn01_8 (2.12a.wav):
*LYN: and the first part of it /=TOP= is like /=INT= oh we have lecture /=CMM= then we have lab //=CMM=

b. afammn03_5 (2.12b.wav):
*ALA: and mom said /=INT= &n [/1]=EMP= you know /=AUX= don't bring it up to Linda /=CMM_r= don't tell her about Mike //=CMM_r=

c. afammn03_28(2.12c.wav):
*LEE: go big /=CMM= or don't go at all> //=CMM=

As an illocutionary unit, CMM has a prosodic nucleus whose form vary according to the type of illocution conveyed. What is deemed fundamental for considering two or more illocutions as belonging to the same pattern is the fact that they convey a clearly recognizable rhetorical effect.

(2.13) afamdl01_78 (ex2.13.wav):
*RIC: I figure /=INT= the more cards I get out /=CMM= the more people I talk to /=CMM= the more cars I'm gonna sell //=CMM=

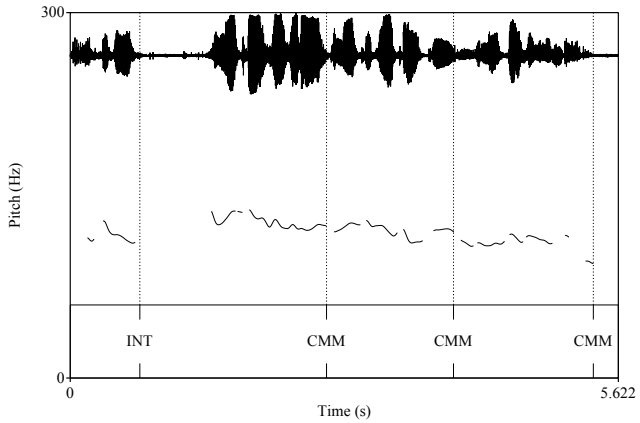


Figure 2.9 – Waveform and *f0* contour of the CMM chain in ex. (2.13)

The CMM units in the utterance above constitute types of assertions that present states of affairs that are correlated with one another. The three CMMs are connected with one another by means of a melodically cohesive pattern. As said above, the compositional illocutionary pattern is not the only case in which an utterance carries more than one illocution. The next section presents the stanza, a pragmatic unit also characterized by the presence of more than one illocution.

2.3.6 The stanza and the bound comment (COB)

A *stanza* may be thought of as a special type of utterance, one which is composed of at least two illocutionary units featuring a prosodic signal of continuity. The illocutionary units in a stanza are called *bound comments* (COB), except for the last one, which keeps the usual COM label, since it is not marked with the continuity signal.

Unlike a chain of CMMs, which supposedly stem from a common internal drive, a sequence of COBs does not impart any particular rhetorical effect. On the contrary, the COB units in a stanza constitute series of weakened and homogeneous illocutions pertaining to the same class (MONEGLIA; RASO, 2014), as the example below shows.

(2.14) afamdl01_9 (ex2.14.wav)

*FRE: and then he comes into the cafeteria /=COB= and I thought he was coming in to chase everybody away /=COB= <but he's coming> after me //=COM=

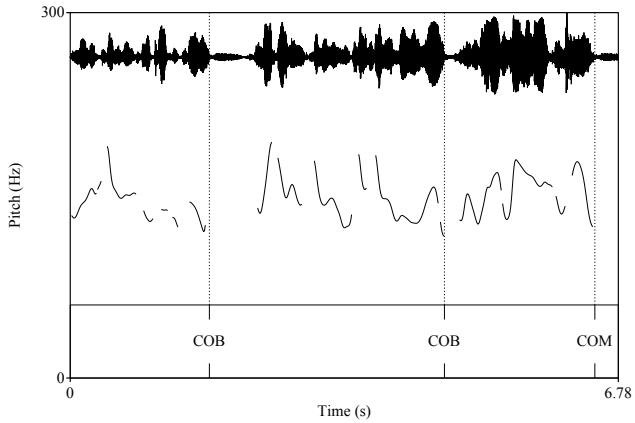


Figure 2.10 – Waveform and *f0* contour of the stanza in ex. (2.14)

The stanza in (2.14) is composed of two COBs and one COM, which show no prosodic pattern but represent the same illocutionary class. In other words, they constitute is a string of assertions that, despite being part of the same sequence marked with a terminal prosodic break, do not constitute a combined set of illocutionary units jointly conveying a rhetorical effect.

Stanzas are typically found in monologues, a type of interaction in which the speaker usually has much more control over the turn as well as more time available for planning the utterances. Thus, narratives, expositions, and the like comprise spoken genres in which stanzas are expected to be frequent.

In addition, the stanza may exhibit different degrees of informational complexity, ranging from relatively simple patterns, as in (2.14) above, to very complex patterns, as the example below shows.

(2.15) afammn01_4 (ex2.15.wav):

*LYN: that's another thing too /=COB= is I kinda had a general idea of [/1]=SCA= of kinda how to do it just watching him /=COB= or [/1]=EMP= and watching /=SCA= people come to our place and doing it and stuff /=COB= you know /=AUX= and /=AUX= I don't know then /=COB= down there /=TOP= &he /=EMP= it's mandatory /=COB= you have to [/3]=EMP= &he /=EMP= to graduate /=TOP= you know /=AUX= or well /=AUX= to /=SCA= get the degree /=TOP= you know /=AUX= you have to take this class //=COM=

2.3.7 Interrupted, scanning, and empty prosodic units

The units of the information and prosodic patterns sometimes do not exhibit a one to one correspondence. For instance, a textual information unit may at times be interposed with another information unit that can be either textual or dialogic. This type of phenomenon has been shown in the section in which the PAR was discussed. The one-to-one correspondence between the intonation and information units is not observed in other cases as well. *Scanning units* (SCAs), for example, constitute a part of a textual information unit, but its function is consigned to carrying morphosyntactic material. Examples (2.16) and (2.17) below illustrate two slightly different phenomena in which SCAs may be involved.

(2.16) afamcv02_8 (ex2.16.wav, ex2.16a.wav, ex2.16b.wav):

*SHR: <hhh> it's /=SCA= roast beef with /=SCA= melted cheese /=COB= and sautéed onions //=COM=

The utterance above contains a COB that is realized by means of three prosodic units, the first two of which being SCAs which, as such, simply carry locutive material. This can be appreciated by listening to ex2.16a.wav and ex2.16b.wav. The first file contains only the SCA units, and the hearer cannot give a pragmatic interpretation to them. File 2.16b.wav, however, is perfectly interpretable, as it contains the part of the information unit that bears the features signaling its information function.

(2.17) afamdl02_9 (ex2.18.wav):
*DAR: <but /=AUX= I> didn't like the book the way I [/3]=SCA= the minute I looked at it //=COM=

The utterance in (2.17) illustrates a case that is slightly different. About halfway through the utterance, the speaker appears either to change the initial program or to utter an unintended word. So, instead of completing the phrase "the way I", she changes it to "the minute I (…)". This is why the slash sign that indicates the non-terminal break is followed by a digit, which is meant to specify the number of words that has been "cancelled" by the speaker. This is phenomenon, referred to as *retraction*, happens considerably often in spoken language. Retractions may be thought of as a "prosodic eraser" whereby speakers make minor adjustments in utterances without having to abandon them completely to start another.

It may also happen that every single word in a given prosodic unit be cancelled, like in (2.18) below. In those cases, the prosodic unit carrying the cancelled words lack both information and morphosyntactic functions, hence the label EMP, which stands for *empty unit*.

(2.18) afammn01_38 (ex2.18.wav, ex2.18a.wav):
*LYN: we did [/2]=EMP= we did /=CMM= a lot of stuff //=CMM=

To be sure, meaningful repetitions do occur in speech. But as ex2.18a.wav shows, that is not the case with the utterance in the example above, in which the first prosodic unit has no function, being just a hesitation or some phenomenon of the like.

Utterance interruption, i.e. when the speaker abandons his speech program by either starting a new utterance or quitting the turn completely, occurs very frequently as well. Before abandoning an utterance, sometimes the speaker finishes the information unit, in which case it is possible to establish the informational nature of the unit where the interruption occurred. If the informational function is unidentifiable, the tag UNC is used. The examples below illustrate the two cases.

(2.19) afamdl04_141-142 (ex2.19.wav):
*JUL: <and> we paid twenty-five hundred for her //=COM= and she was a nine-year-old Anglo Arab /=COB= &he /=TMT= about /=INT= fifteen three +=COB=

(2.20) apubdl01_6-7 (ex2.20.wav)
*RAN: if you just [/3]=EMP= if you just +=UNC= all you have to do is stand up /=COB= and look out the window /=COB= and look at the bright /=COB= and talk //=COM=

Finally, example (2.21) shows a case of time-taking (filled pause), which involves sustaining a certain vocalic sound for a moment. Although the quality of the vowel-like sound changes depending on the language and culture of the speaker, it is a phenomenon common across languages. The label TMT is used to annotate these units, which have no informational function properly speaking.

(2.21) afamdl04_135 (ex2.21.wav)
*JUL: like with /=SCA= &he /=TMT= yyy's horse /=TOP= I looked at twenty-eight horses in ten <days> /=COB= before I found the one that <I thought would suit> her //=COM=

In the next section, the dialogic information units are briefly described.

2.3.8 Dialogic information units

Dialogic information units are equivalent to what is referred to as discourse markers within other frameworks (see RASO, 2014, RASO; VIEIRA, 2016, GOBBO, 2019). The function of dialogic units is associated with the regulation of the interaction. They may be used for numerous purposes, including to push the interlocutor to do or stop doing something, to maintain the communication channel open, to establish a link between two utterances, among other things.

Traditionally, L-AcT identifies six types of dialogic units, namely the incipit, the phatic, the expressive, the allocutive, the conative, and the discourse connector. In the AE minicorpus, dialogic units have been mostly annotated with the tag AUX. This is because these units require a level of acoustic quality that the AE minicorpus frequently lacks (see chapter 4).

As dialogic units are not of primary concern in this study, they will be described more succinctly. For examples and further details, see the references mentioned in the beginning of this section.

The *incipit* (INP) has the function of initiating the utterance or a subpattern in a stanza, establishing an affective contrast with the utterance or stanza sub-pattern that precedes it. Regarding its prosodic features, INP shows high rates of *f0* variation, high intensity values, and short duration. It always occurs in the beginning of utterances or stanza sub-patterns.

The *phatic* (PHA) is a unit that, freely distributed inside the utterance, helps maintain the communication channel open. It exhibits flat or falling *f0* profile, very short duration and low intensity values. Frequently, its locutive material is phonetically reduced, as result of its typically high articulation rates. When a phonetic reduction occurs, the locutive material may only be distinguished if heard together with the surrounding prosodic units.

The *expressive* (EXP) supplies emotional support for the illocution, possibly establishing social cohesion with the addressee. In addition, it can take part in turn-taking moves, appearing in the beginning of utterances but, unlike INP, establishing no contrast with the preceding discourse. Although EXP may show different prosodic profiles, which varies according to the particular role it plays, it frequently shows some modulation and intermediate duration and intensity values.

The *allocutive* (ALL) identifies the individual to whom the utterance is addressed and also marks social cohesion. It must not be confounded with *vocative*, a notion which conflates (i) addressee identification and (ii) the illocutionary force of calling (see RASO, 2014, p. 424-26). ALL has the function described in (i), and it cannot stand alone on its own – hence, it

cannot be an illocution. The function described in (ii) can only be performed by an illocutionary information unit, which may be COM, CMM, or COB, depending on how it is uttered.

ALL shows a flat or falling prosodic profile, low intensity values, and short duration, except when it occurs at the end of the utterance, in which case it usually exhibits some degree of syllable lengthening. As for distribution, ALL may occur at any position inside the utterance, but different languages and cultures seem to favor particular positions.

The *conative* (CNT) has the function of pushing the addressee to start or stop doing something; therefore, the occurrence of this unit seems to correlate with the occurrence of directive illocutions (see MONEGLIA, 2011, ROCHA, 2016). Although more commonly found at the beginning of utterances, CNT has free distribution. It has a falling or modulated profile, short duration and high intensity.

The *discourse connector* (DCT) functions as a link between two utterances or stanza sub-patterns, marking continuity without signaling contrast between the units it links. Adverbs, conjunctions and even prepositions are common morphosyntactic fillers in DCTs. They are typical of monologues or passages containing narrations, explanations and argumentations. DCT shows flat, smoothly rising or falling profile, long duration, high intensity and low *f0* variation rate.

In summary, dialogic units do not partake in the makeup of the syntactic and semantic content of the utterance. On the contrary, they are involved in regulating the interaction and are often directed towards the addressee. The morphosyntactic filling of a dialogic unit is never compositional with locutive content of neighboring units. To be part of a dialogic unit, a linguistic expression must be realized in a dedicated prosodic unit which cannot be pragmatically interpreted as autonomous units.

2.4 Final words

In this chapter, some of the main concepts and terms of the framework that is followed in this dissertation were laid down. The main goals were to specify the theoretical underpinnings on which this study is based, thus providing a background for the discussion in the following chapter on the ideas related to the term *topic*.

3 The notions of topic

The notions associated with *topic* and *comment*, as well as the more or less equivalent terms *theme* and *rheme*, vary considerably depending on the approach considered. In general, the definitions found in the literature are based on perspectives centered on syntactic and semantic notions like *sentence*, *clause*, *propositional content* and *truth conditions*. In addition, other pairs of notions such as *presupposition* and *focus*, *given* and *new*, and *thetic* and *categoric* are sometimes equated – or at least correlated with – topic and comment, which suggests a mix-up between linguistic and cognitive categories.

Within some approaches to information structure, restrictions are imposed on the cognitive status of a referent evoked by a topic expression. For instance, it is often required that, in order to qualify as topic, a linguistic expression must refer to something which at least recoverable from the context and be realized as a noun phrase occurring at the beginning of the sentence. Other frameworks, however, establish no formal constraint for an entity or expression to be topic. In fact, topic can be a notion understood in terms of a pragmatic relation established between mental representations that do not even have to be instantiated by a linguistic expression at all.

Different authors[20] trace the history of the notions associated with topic back to Aristotle's *Organon*, and particularly to *Categories* (see BARNES, 1984), where the philosopher discusses the concepts of *subject* (i.e. an underlying thing or being) and *predicate* (i.e. the thing that is said about the subject) and the relationships that can be established between them. As pointed out in chapter 1, Weil's (1978 [1844]) *point of departure* and *enunciation* is also relevant for the historical developments associated with the term under discussion, and it must be pointed out that the Arabic tradition has contributed with similar notions as well, namely *ibtidāʾ* (equivalent to topic or theme) *mubtadaʾ* (equivalent to comment or rheme), discussed in Marogy (2010, p. 151-99) and mentioned in Krifka (2008).

The term *theme* predates the term *topic*. In its original formulation, *theme* describes the portion of the sentence that contains "things relatively familiar or most readily available to the speaker as the starting point" (MATHESIUS, 1983[1929], p. 127). The unmarked theme, according to Mathesius's perspective, comes first in the sentence and is followed by the *rheme*, which "expresses something relatively new and contains what is asserted by the sentence"

[20] Krifka and Musan (2012), Marogy (2010), Ping (2005), Lambrecht (1994).

(MATHESIUS, 1983[1929], p. 126). As for the term topic, it was introduced by Hockett (1958), who wrote that

> [t]he most general characterization of predicative constructions is suggested by the terms 'topic' and 'comment' for their ICs [immediate constituents]: the speaker announces a topic and then says something about it (HOCKETT, 1958, p. 201).

In this frequently cited passage, Hockett inaugurates the use of a term which has been embraced by linguists of different persuasions.

Going back to theme and rheme, they constitute alternative terms to *psychological subject* and *psychological predicate* (GABELENZ, 1869), which Mathesius deemed inappropriate, on the grounds that they may be confused with the grammatical sense associated with subject and predicate. In pointing to this confusion, Mathesius reveals a certain concern in differentiating what is cognitive from what is linguistic, which is something that does not seem to be particularly common in the literature on information structure (IS).

Halliday (1967b), building on insights by Mathesius and other researchers from the Prague School, distinguishes between theme and topic, considering the latter a subtype of the former. He defines *theme* as the starting point of the clause on the message level corresponding to the constituent that occurs the initial position in the clause. As for the term *topic*, Halliday objects that it "tends to be used as a cover term for two concepts that are functionally distinct, one being that of Theme and the other being that of Given" (HALLIDAY; MATTHIESSEN, 2014[1985]). As this passage shows, Halliday keeps theme and givenness separate. In fact, in his systemic approach, givenness belongs to a separate level that runs parallel to the thematic system.

In the remainder of this chapter, different definitions and ideas related to the term topic in the linguistic literature will be reviewed. Similar to what happens with other IS terms (see chapter 1), linguists tend to have particular views as to what the term topic describes. In part, this is to be expected, given the abstract nature of IS notions and the problems involved in correlating language form with cognitive and pragmatic categories.

The aim of this chapter is to provide a theoretically diverse background for an adequate assessment of the notion of topic (TOP) within the L-AcT framework, which, as shown in the previous chapter, defines topic as an information unit which, having specific prosodic properties, supplies an identification domain for the interpretation of the illocution. The abbreviation "TOP", as said previously, will be used to refer to the concept of topic according to L-AcT. All other notions will be referred to as "topic". We will begin reviewing the widely referenced paper on topic typology by Li and Thompson (1976).

3.1 Topic according to Li and Thompson

Li and Thompson (1976) propose a classification of languages based on a continuum that goes from topic-prominent languages to subject-prominent languages. In between these extremes, there are languages that exhibit both types of prominence and languages for which neither topic nor subject are appropriate descriptive categories. The examples below[21] show the difference between the two types of articulation.

(3.1)

a. $[\text{John}]_{\text{subject}}$ $[\text{hit Mary}]_{\text{predicate}}$

b. $[\text{As for education}]_{\text{topic}}$, $[\text{John prefers Bertrand Russell's ideas}]_{\text{comment}}$

The parameters that, according to Li and Thompson, distinguish subjects from topics are related to both semantic and syntactic aspects, as seen in Table 3.1.

Table 3.1: Topic and subject according to Li and Thompson (1976, p. 461-466)

Parameters	**Topic**	**Subject**
Definiteness	always definite	–
Selectional relations	–	an argument of the predicative structure
Predictability	unpredictable	partially predictable, as it is an argument
Function	framework for the predication	varied; may not even have one (e.g. *dummy subject*)
Verb agreement	rare	always
Linear distribution	beginning of the sentence	free
Grammatical processes	None, since syntactically independent	reflexivization, passivization, imperativization etc.

Based on these parameters, Li and Thompson (1976) propose that languages can be placed in one of the four groups listed below, depending on the relevance of topic and/or subject in the language in question. The authors exemplify each group as follows.

a. subject-prominent languages: sentences are best described in terms of subject-predicate structures, e.g. Indo-European, Niger-Congo, Finno-Uralic, Semitic, and Indonesian families

[21] Li and Thompson (1976, p. 459).

b. topic-prominent languages: sentences are best described in terms of topic-comment articulation, e.g. Chinese, and Lahu and Lisu
c. topic and subject languages, in which the two are equally prominent, e.g. Japanese and Korean
d. no prominence of either topic or subject, e.g. Tagalog and Illocano.

Li and Thompson provide no clear prosodic specification for topic, and what appears to be the most relevant difference between topic and subject is that the former is not part of the argument structure of the sentence, but the latter is. In fact, their definition is similar to that proposed in Chafe (1976), a paper published in the same volume as Li and Thomson (1976). Chafe's view is outlined in the next section.

3.2 Topic according to Chafe

In the same volume where Li and Thompson (1976) propose their typology, Chafe defines *topic* as "the frame within which the sentence holds" (CHAFE, 1976, p. 51). Chafe saves the notion of *aboutness* for his definition of *subject*, which he describes as the constituent in the sentence that specifies a cognitive domain to which information is contributed. As can be seen, Chafe's subject is very similar to Hockett's (1958) topic. In line with Li and Thompson, for Chafe the topic constituent does not participate in the sentence as an argument, whereas subjects always do.

Chafe distinguishes topic, a linguistic category, from givenness and contrastiveness, which pertain to the cognitive level. His notion of givenness is based on the perspective of the speaker. So, for instance, *given* information is what the speaker considers to be present or activated in the hearer's consciousness. With regards to form, given information is usually associated with attenuated prosody – e.g. constituents that carry given information typically have short duration and lower pitch – as well as pronominalization and ellipsis. Topic-prominent languages such as Mandarin (see Li and Thompson's typology above), according to Chafe, provide prototypical examples for his notion of topic.

(3.2) [nèi-xie shùmu] TOPIC shù-shēn dà[22]

[22] Chafe (1976, p. 50).

[those tree] tree-trunk big

The constituent “nèi-xie shùmu” – the topic in (3.2) – has the function of restricting the domain of the predication expressed in the remaining part of the utterance. That is, this topic establishes that “the property of being big” is being assigned not to any “tree trunks”, but precisely to the trunks of the trees that the topic expression “nèi-xie shùmu” denotes.

As for *contrastiveness*, it is associated with the presence of alternatives. According to Chafe, constructions like “as for those trees, their trunks are big” do not constitute appropriate translations for the Mandarin sentence in (3.2). This is because English constructions of this type imply contrast – i.e. the presence of alternatives – which is absent in the Mandarin example. Analogously, a construction such as “the PLAY/as for the PLAY, John saw it yesterday” (CHAFE, 1976, p. 49, adapted), where uppercase letters denote prosodic prominence, would not constitute a good example of topic but rather of contrast.

The prominence signaled by uppercase letters in the example above is interpreted by Chafe as an indication of contrast, which tells the addressee that there is at least one denotation alternative that is compatible with what the expression “the play” refers to. But what exactly is the nature of this prominence? Is it signaled by f0 variation, by syllable lengthening, or perhaps some combination of different prosodic parameters? This imprecision – understandable given the stage of technology development at the time Chafe (1976) was published – still persists in the IS literature.

As argued in chapter 2, prosodic prominences may carry out very different functions. Some prosodic prominences have the ability to impart the illocutionary force. There is also that kind of prosodic prominence which communicates that the content of a prosodic unit should be interpreted as a domain for the interpretation of the illocution. In any case, one can only determine the function of a given prosodic prominence by listening to it.

The prominence in “the play”, to be sure, could indeed communicate a contrast of some sort. But because we do not know exactly what type of prominence, we cannot be sure about the function Chafe is associating it with. As a matter of fact, this issue, which in the past is justified by the lack of technological means, still persists in the IS literature.

As in the Mandarin construction in (3.2), certain adverbs and prepositional phrases in English can, according to Chafe, behave like frames for the interpretation of the predication. For instance, in the sentence “In Dwinelle Hall people are always getting lost”, the phrase “in Dwinelle Hall” restricts the domain within which the predication applies (CHAFE, 1976, p.

51). However, unlike the Mandarin construction, it requires the preposition. Note that, according to this account, "people" is what establishes what the sentence is *about*, while "in Dwinelle Hall" is interpreted as the domain within which the proposition conveyed by the sentence holds.

In the next section, another influential notion of topic, namely, Reinhart's (1981, 1982) aboutness-topic will be reviewed.

3.3 Topic according to Reinhart

Reinhart's contribution in the area of IS studies is widely acknowledged (see KRIFKA, 2008, PETROVA; SOLF, 2009). She regards *topic* basically as the Anglo-American equivalent of Mathesius's theme, and she centers her discussion on the notion of sentence topic, which is restricted to referential NPs in declarative sentences used as assertions.

Sentence topic is defined as "the expression whose referent the sentence is about" (REINHART, 1981, p. 57). In her account, the topic must be linguistically encoded in the sentence but not necessarily as an initial expression. Drawing on the notion of *context set*, which following Stalnaker (1978) is understood as the set of propositions accepted as true at the time of discourse, she proposes that a sentence topic is

> one of the means available in the language to organize, or classify the information exchanged in linguistic communication – they are signals for how to construct the context set, or under which entries to classify the new proposition (REINHART, 1981, p. 80).

Reinhart's topic is like the heading of a file card (HEIM, 1982, 1983) – an analogy eventually taken up by other authors – corresponding to an instruction to the hearer as to under which entry to enter and store the information conveyed by the sentence. It is an approach built on ideas which assume an algorithmic model of cognition, relying on notions like mental representation and propositional meaning, providing no room for a proper treatment of the action dimension of language. In the next section, a related approach to Reinhart's will be reviewed.

3.4 Topic according to Vallduvì

In Vallduvì's account of IS in sentences, he uses the term *link* to describe something admittedly very similar to Reinhart's sentence topic:

> A link is *an address pointer* in the sense that it directs the hearer to a given address (or file card in Reinhart's (1982) or Heim's (1983) terms) in the hearer's knowledge-store, under which the information carried by the sentence is entered (VALLDUVÌ, 1992, p. 59, emphasis added).

But the link, unlike Reinhart's sentence topic, is always a sentence-initial expression. It is an optional part in the hierarchical structure represented in (3.3).

(3.3) S = [FOCUS, GROUND(LINK, TAIL)][23]

The general informational structure for sentences, according to Vallduvì, is represented in (3.3): the mandatory focus, and the optional link and tail, which are elements of the ground. Vallduvì's notion of *focus* describes the portion of the sentence marked with a prosodic prominence of intonational nature which encodes relevant information. In other words, the focus of a sentence corresponds to a knowledge contribution to the hearer (VALLDUVÌ, 1992, p. 198).

The *ground*, on the other hand, constitutes the vehicle for the focus. It is, therefore, an instructional frame that allows the hearer to adequately enter the information in her "knowledge file". The information encoded in the ground is knowledge that the speaker assumes to be already present to the hearer, and Vallduvì claims that the ground has nothing to do with logico-semantic meaning. Also, a sentence may have no ground whatsoever if the speaker considers that the addressee can figure out by herself how the sentence contributes to her knowledge.

Besides connecting the information with a cognitive entity by means of the link, the ground may be accompanied by a *tail*, "negatively identified as the nonfocal nonlink part of the sentence" (VALLDUVÌ, 1992, p. 61). It supposedly aids the link in providing additional instructions as to how the information must be entered by the hearer.

According to this account, there are four ways in which a sentence can be organized informationally: (i) link-focus, (ii) all-focus, (iii) link-focus-tail, and (iv) focus-tail. Some of these structures can be exemplified as follows.

(3.4)[24]

a. [The boss]$_{link}$ [CALLED]$_{focus}$

b. [The boss]$_{link}$ [I wouldn't BOTHER]$_{focus}$

---

[23] Vallduvì (1992, p. 57, adapted).
[24] Vallduvì (1992, p. 62-4).

c. [Broccoli]$_{link}$ [the boss]$_{link}$ [doesn't EAT]$_{focus}$

d. [The boss]$_{link}$ [HATES]$_{focus}$ [broccoli]$_{tail}$

As seen in (3.4), the link can be either the subject, as in (3.4a) and (3.4d), or some other fronted expression, as in (3.4b) and (3.4c). As for the tail, the constructions in which it occurs, according to Vallduvì, correspond to what is referred to in the literature as a narrow-focus construction, i.e. a construction in which the prosodic prominence underscores, so to speak, a specific element within a syntactic phrase. As part of the ground, the content of the tail only serves to guide the hearer in her entering the information conveyed by the focus.

As can be seen from this brief outline, this approach is centered on the sentence. The link, Vallduvì's equivalent for topic, is regarded in terms of an instruction. The type of instruction provided, however, is one that has nothing to do with the illocutionary dimension, and the underlying motivation of the approach in question is to treat IS from computational viewpoint.

3.5 Topic according to Lambrecht

According to Lambrecht (1994), *topic* is a pragmatic notion defined as "the thing [the referent] which the proposition expressed by the sentence is about" (LAMBRECHT, 1994, p. 118). Drawing on the *Principle of Relevance* (STRAWSON, 1964), Lambrecht identifies the topic of a sentence as the element that constitutes the matter of interest and in relation to which the proposition conveyed by the sentence can be said to be relevant.

As this definition suggests, identifying the topic of a sentence is not always a trivial matter, and it may be the case that no constituent in the sentence corresponds to its topic. Ultimately, identifying the topic is something that can only be done if the context is taken into account. By *context*, Lambrecht means the communicative intentions of the speaker and the state of mind of the hearer. For example, the sentence "the children went to school" cannot be said to convey information about "the children" unless the proposition it conveys can be *pragmatically construed* as being about "the children".

Pragmatic construal has to do with how the proposition is being added to the hearer's mind. Thus, "the children went to school" may be said to be about "the children" if, for instance, it is uttered in response to a question like "what did the children do?". This is because in such circumstance, the sentence can be construed as an act on the part of the speaker to increase the hearer's knowledge about "the children". Nevertheless, "the children went to school" may be

uttered in response to a question such as “is it okay for us to eat the ice cream now?”, in which case it would not be congruent to construe the proposition carried by “the children went to school” as being about “the children”.

Lambrecht claims that, in languages like English, sentences exhibiting a topic-comment structure usually have an accented element in the verb phrase as well as a low pitch prominence in the topic constituent, which, according to his account, is always realized by an NP. So “the children” in the sentence above, assuming it is realized as topic in a topic-comment sentence, would likely exhibit a low pitch prominence while “school” would be accented. In addition, topic-comment sentences are taken to be prosodically unmarked with respect to information structure, which means that their prosodic characteristics are also compatible with structures in which the subject is not topic (LAMBRECHT, 1994, p. 122).

What appears to be one of the most relevant aspects about Lambrecht account of topic is that, like the approaches reviewed so far, topic is defined in terms of mental representations, based on a model of language that is centered on the notion of sentence. It will be shown below (see also chapter 2) that L-AcT provides an interesting alternative to this view, as this framework defines TOP as a notion inextricably related to the illocution, rather than propositional content. In addition, Lambrecht’s topic is something hard to be treated linguistically, given the frequent lack of formal markings associated with it and its substantial contextual dependence.

In the next section we present a definition of topic that is quite distant from the one adopted in this dissertation and also from those reviewed so far.

3.6 Topic according to Givón

Givón uses the term *topic* to describe the degree of importance of participants in a given piece of discourse. In his account of what he refers to as *topicality*, reference has a special significance. The function of topicality has to do with two pragmatic aspects of reference: (i) referential accessibility and (ii) thematic importance, which constitute, respectively, the anaphoric and cataphoric facets of reference (GIVÓN, 2001b [1984], p. 227).

Referential accessibility is involved in the speaker’s choice for a specific form based on her judgements about how accessible a referent is for the addressee. The anaphoric nature of this aspect of topicality is manifested in the idea that these judgements are made on the basis of the preceding discourse.

Thematic importance, on the other hand, is related to the instructions given by the speaker as to the attentional activation of a referent in the subsequent discourse. Givón writes that

> [t]he propositional information coded in state/event clauses is typically about some topical participant(s) in the state/event. Such topical participants are most commonly the subject, direct-object or indirect-object of the clause, and thus most commonly noun phrases ('entities') rather than verbs ('events') or adjectives ('states') (GIVÓN, 2001b, p. 253-54).

The examples below[25] illustrate what Givón understands as topic and how it may be realized grammatically.

(3.5)

Context: What did you do with the paint?

a. I sprayed *it* on the wall.

b. ?I sprayed the wall with *it*.

(3.6)

Context: What did you do to the wall?

a. I sprayed *it* with paint.

b. ?I sprayed the paint on *it*.

The context in (3.5) establishes the referent designated by "the paint" as the topic. As an available referent, it can then be picked up by the pronouns in (3.5a) and (3.5b). The situation is quite the same in (3.6), with the difference that the topical referent now is established by "the wall" instead.

Givón talks about certain grammatical functions as being more topical then others, with subjects being the grammatical function with the highest degree of topicality, followed by direct object and indirect object. This hypothesis, Givón claims, accounts for the fact that replies in (3.5a) and (3.6a) sound more natural than their counterparts in (3.5b) and (3.6b).

Givón proposes that certain grammatical constructions are specialized in signaling topical referents that exhibit a relatively low accessibility. Low accessibility is associated with a particular referent being (i) introduced for the first time, (ii) reintroduced after being deactivated

---

[25] Givón (2001b, p. 148). The superscript "?" sign denotes questionable appropriateness.

for some reason, and (iii) potentially ambiguous as a result of referential competition or contrast.

Givón's topic is very unique when compared to the other notions of topic reviewed in this chapter. It constitutes a more discourse-oriented definition of the term, and it is reviewed here not only because it is widely cited in the IS literature, but also because one of the aims of this chapter is to illustrate how varied and therefore confusing the theorization in the area is.

3.7 Topic according to Erteschik-Shir

Erteschik-Shir (2007, p. 19) associates *topic* with (i) what the sentence is about, (ii) given information, and (iii) pivot for truth value assessment. Previously mentioned elements, the author says, make for potential topics, but an element simply has to be contextually available or somehow susceptible to accommodation in order for it to qualify as a potential topic. Regarding (iii), she considers that a DP without reference in topic leads to a truth-value gap, which would come as a corollary of the notion that the topic is what the sentence is about, after all, a sentence cannot be said to be about nothing. This is a perspective that she draws from Strawson (1964).

Erteschik-Shir argues that topic and focus, the two primitives of her IS model, can be marked by means of intonation, word order and morphology. According to her, topic and focus constitute triggering instructions for the handling of referents in discourse. Based on Heim's (1982) model of common ground (CG) in terms of file (see chapter 1), Erteschik-Shir formulates the role of these notions as follows[26]:

- *Topic*: instructs the hearer to locate on the top of his file an existing card with the appropriate reference
- *Focus*: instructs the hearer either to create a new card (a cognitive domain) and put it on the top of the file (i.e. the forefront of the CG) or locate an existing card and put it on the top of the file.

The sentences in (3.7) below, particularly the answers, provide examples of what Erteschik-Shir considers to be topic.

---

[26] Erteschik-Shir (2007, p. 44-5).

(3.7)[27]

a. What did John do?
a'. he $_{top}$ [washed the dishes] $_{foc}$

b. What did John wash?
b'. he $_{top}$ washed [the dishes] $_{foc}$

c. Who washed the dishes?
c'. John $_{foc}$ washed them $_{top}$

d. What happened to the dishes?
d'. [John washed them $_{top}$] $_{foc}$

e. What did John do with the dishes?
e'. [he $_{top}$ [washed them $_{top}$] $_{foc}$] $_{foc}$

f. What happened?
f'. $_{s}$ top $_{t}$ [John washed the dishes] $_{foc}$

The sentences in (3.7a), (3.7b), and (3.7f) above illustrate what the author considers to be unmarked topics, i.e. topics that precede the focus. The difference among them is that in (3.7a) and (3.7b) the topics are realized by subjects, whereas in (3.7c) the topic is the here-and-now of the utterance, or what she refers to as *stage topic* (the subscribed "s" and "t" stand for "space and "time").

In (3.7c) and (3.7d), the object pronoun "them" – or rather its referent – is the topic. The difference between these two examples is that in (3.7d) the topic is embedded inside the focus. On the grounds that both topics and foci should be realized by syntactic constituents, Erteschik-Shir argues that considering the topic as part of the focus is better than considering the focus to encompass only "John washed", which would map the focus to a sequence that is not a syntactic constituent.

As for the sentence in (3.7e), Erteschik-Shir posits two superimposed structures in it. On one level, "he" is the topic and "washed them" is the focus; on the other level, the entire sentence is focal, since it is an all-new or thetic sentence. She assumes that the topic-focus level, considered unmarked, takes precedence over the all-focus level, and that the topic-focus structure constitutes the domain in which truth-value assessment takes place.

---

[27] Erteschik-Shir (2007, p.1-3).

3.8 Krifka

Krifka (2008) also regards topic in terms of semantic aboutness: “the topic constituent identifies the entity or set of entities under which the information expressed in the comment constituent should be stored in the common ground content” (KRIFKA; MUSAN, 2012, p. 38). It is interesting that, apart from the term “constituent”, this definition does not contain any formal specifications.

Drawing on Reinhart (1981) and Heim (1982), Krifka compares the topic of a sentence to the heading of a file card on which the information conveyed by the comment should be entered and stored.

(3.8)

a. [Aristotle Onassis]$_{\text{topic}}$ [married Jacqueline Kennedy]$_{\text{comment}}$

b. [Jacqueline Kennedy]$_{\text{topic}}$ [married Aristotle Onassis]$_{\text{comment}}$

The utterances above, which convey the same propositional content, contain different topics, i.e. different files on which to add the proposition. In (3.8a), the proposition should be added to the file card identified as “Aristotle Onassis”, while in (3.8b) it should be added to the one identified as “Jacqueline Kennedy”. Krifka’s topic is not necessarily given – i.e. previously mentioned or contextually inferable for some other reason –, though he states that given topics are much more frequent than new ones. The sentence below is provided as an example of new topic.

(3.9) [A good friend of mine]$_{\text{topic}}$ [married Britney Spears last year]$_{\text{comment}}$

Regarding the sentence above, Krifka and Musan (2012) argue that not only does the topic identify the entity which the comment is about but also introduce a new file card to the common ground. This is a job that, in Erteschik-Shir’s (2007) model, is done not by the topic, but rather by the focus (see section 3.7).

Krifka and Musan (2012) also say that there may be utterances with more than one topic and even without topic. In the first case, the authors refer to sentences whose topic constituents identify more than one entity in relation to which the information in the comment should be added to the common ground. For instance, in the sentence “as for Jack and Jill, they got married

last year", the topic constituent "as for Jack and Jill" indicates that the proposition conveyed is about Jack and Jill taken individually. Krifka and Musan (2012) suggest alternatively that it is possible to consider that "Jack and Jill" identify one single card with two relevant entities.

Regarding sentences without topics, Krifka and Musan (2012) say that a sentence lacking a topic constituent does not necessarily lack a domain of relevance. This is because they distinguish between *topic constituent* and *topic denotatum*. The all-focus sentence "the HOUSE is on FIRE" – considered as such because of the double prosodic prominence" – would thus have a topic denotatum possibly identified as contextually accessible situation.

Krifka and Musan (2012) also talk about a category that melds aboutness topic and focus and is called *contrastive topic*. As the term itself suggests, this category is defined as a topic constituent which, by means of a particular intonation, does the additional job of indicating alternatives, like the example below shows.

(3.10)[28]

a. What do your siblings do?

b.[My [SISter]$_{focus}$]$_{topic}$ [studies MEDicine]$_{focus}$ and [my [BROther]$_{focus}$]$_{topic}$ [is working on a fright ship]$_{focus}$

In addition, a contrastive topic would suggest that the current sentence does not satisfy the expected information entirely, and it may be the case that the contrastive topic indicates that contributed information and expected information do not mesh.

(3.11)[29]

a. Does your sister speak Portuguese?

b. [My [BROTHer]$_{focus}$]$_{topic}$ [DOES]$_{focus}$

The answer in (3.11b) does not satisfy the question in (3.11a), which requests a specific piece of information regarding a person in particular, namely the addressee's sister. The reply in (3.11b) a sort of rectification, and the contrastive topic "my brother" could be interpreted as signaling that. Thus, the authors conclude that, while it is often the case that contrastive topics indicate alternatives, they do not have to do so.

---

[28] Krifka and Musan (2012, p. 30-31).
[29] Ibidem.

An additional notion related to topic in Krifka's approach is the case of *frame setters*, which establish a frame for the interpretation of the expression that follows it.

(3.12)[30]

a. How is John?

b. Healthwise, he is [FINE]$_{focus}$

The expression "healthwise" in (3.12b) does the job of specifying the domain to which the predicate "fine" applies: "he" is fine in terms of health, but may not be so financially, for example. According to the authors, instead of regarding frame setters strictly in relation to the predication – as Chafe (1976) does –, it seems more accurate to say that these elements stipulate the general type of information that the following expression is to provide. This is because there are cases in which the frame-setting expression bears no direct relation to the predication, like (3.13) below, which is considered to be a clearer instance of such function than (3.12) is.

(3.13) As for his health situation, he had a bypass operation recently

Regarding the encoding of topics, the authors mention that English topics carry a fall-rise intonation and sometimes – in English as well as other languages – are realized in a separate intonation unit. Other formal features of topics would include their placement at the beginning of sentences and the possibility of their being signaled by cleft and pseudo-cleft constructions.

Despite the interesting theoretical insights provided by the approaches outlined so far in this chapter, only L-AcT seems to succeed in providing a rigorous definition of topic that combines form and function in a sufficiently clear way to enable a reliable individualization of units in spontaneous speech. Before exploring the notion of topic according to L-AcT, let us briefly examine a recently published experimental study on the topic-comment relation.

3.9 Topic and comment – production and comprehension experiments

Silva and Fonseca (2018) present a study on the comprehension and production of topic-comment and subject-predicate structures in Brazilian Portuguese, focusing on the role of

[30] Ibidem.

prosody in these processes. Mostly adhering to Li and Thompson's (1976) notion of topic, Silva and Fonseca describe the topic-comment structure as a construction in which the internal argument of a verb is realized at the left periphery of the sentence, as the following examples, provided by the authors, show.

(3.14)

a. [The red backpack]$_{topic}$ [Ana bought (it) in a shopping mall]$_{comment}$

b. [The girl]$_{topic}$ [the aunt took (her) to the shopping mall]$_{comment}$

Silva and Fonseca (2018) also assume that this type of construction exhibits particular prosodic characteristics, mainly because it tends to be divided into two intonational phrases (IP)[31], as opposed to subject-predicate constructions, which are typically realized as one IP. In addition to that, they consider that a topicalized element announces what the sentence about, while the comment provides information about it.

The three experiments carried out in their study (ABX discrimination task, cross-modal naming, and self-paced reading and listening tasks) suggest that Brazilian Portuguese speakers are able to distinguish the prosodic characteristics of a topic-comment construction from that of a subject-predicate construction. Furthermore, the authors show that speakers actually resort to prosodic cues in order to distinguish the two structures, in addition to providing evidence that speakers use both syntactic and prosodic cues for the understanding and processing of topic-comment sentences.

3.10 Topic according to L-AcT

Identifying the topic in an utterance is often a problematic task. This is because the notion is frequently defined in functional terms, with stable formal correlates not being usually established. Unsurprisingly, when the notion is defined primarily in terms of contextual parameters, formal indices may not even apply.

According to L-AcT, TOP is a pragmatic notion that is well established in both functional and formal terms[32]. The breadth of the definition as well as the theoretical assumptions that underlie it seem to really set L-AcT apart from other approaches. Functionally, TOP is a unit

[31] IP according to the *Prosodic Hierarchy* by Nespor and Vogel (2007).
[32] For details about how TOP is defined within L-AcT, see chapter 2 and, particularly, chapters 4, 5, and 6.

the establishes a cognitive domain for the interpretation of the illocutionary force, which is carried by the comment (COM). According to L-AcT, COM is the core information unit, and it can even constitute an utterance on its own.

TOP maintains a relation of pragmatic aboutness with COM. The domain that TOP supplies can be spatial, individual, temporal or of some other type. In absence of a TOP unit, the illocutionary force is interpreted according to some contextually given domain. Also, there is no requirement regarding the information status of the referent denoted by the expression in TOP, since in order for an expression to eligible for filling a TOP unit, it just has to be referential. Like any other information unit defined by L-AcT, TOP does not have to be realized by a specific syntactic constituent. On average, 43% of TOPs are composed of a NP and 33% of TOPs are composed of a VP. The remaining 24% of TOPs are either composed of a PP or an AP. These proportions refer to Italian, Brazilian Portuguese, and American English (see CAVALCANTE, 2015).

TOP is always realized in a position that precedes COM. Furthermore, TOP corresponds to a prosodic unit with specific melodic movements, syllable duration and intensity pattern. The prosodic forms of TOP are carefully examined in chapters 5 and 6, where we present an analysis of TOP units extracted from corpora of different languages and show the consistency of the formal realization of the TOP in spontaneous speech.

The aboutness relation established between TOP and COM is a relation that links a cognitive domain to an illocution. Thus, as Cresti and Moneglia (2011) argue, there is no need to reconduct the TOP-COM relation to a propositional form. In fact, they claim that it would be a mistake to do so. A TOP unit is overtly signaled by the speaker, and its identification is independent of contextual parameters – there is no need to devise a question or to look for a domain in the context in order to identify a TOP unit in an utterance. The reason this is mentioned is that, for many of the authors reviewed in this chapter, the notion of context is very restricted, and frequently modeled as a question.

In the following chapter, we discuss an interrater agreement conducted with four participants to assess the reliability of identification of TOP using the parameters established by L-AcT. In this test we reached kappa coefficients that indicate substantial agreement among the annotators, which speaks to the reliability of the definition and the objectivity of the concept.

In the remaining chapters of this dissertation, we will examine the prosodic features of TOP following L-AcT. We will first deal with some general methodological aspects regarding the corpora used in the study. Then, we will discuss the kappa test that was conducted. In

chapters 5 and 6, we delve into aspects related to the melodic shape and the duration pattern of the unit.

4 The corpora

This dissertation deals with information structure (IS) in a crosslinguistic perspective, following the framework of the *Language into Act Theory* (L-AcT; see chapter 3). The data analyzed here come from four different corpora, all of which belonging to the C-ORAL family. More specifically, the data come from:

- the American English (AE) minicorpus (CAVALCANTE; RAMOS, 2016)
- the Italian (IT) minicorpus[33] (CRESTI; RASO, 2012)
- the Brazilian Portuguese (BP) minicorpus (RASO; MITTMANN, 2011, RASO; CRESTI, 2012)
- the European Portuguese (EP) component of the C-ORAL-ROM corpus (NASCIMENTO et al., 2005), and
- the formal part of the C-ORAL-BRASIL corpus (RASO et al., forthcoming)

The AE minicorpus consists of texts sampled from the *Santa Barbara Corpus of Spoken American English* (SBCSAE; DU BOIS et al., 2000-2005) and was created in accordance with the methodological standards established by L-AcT. The IT has two versions, which are both comprised of texts from the informal Italian component of the C-ORAL-ROM multilanguage corpus (CRESTI; MONEGLIA, 2005). The C-ORAL-ROM documents, besides Italian, other three of the main Romance languages spoken in Europe, namely French, Spanish, and Portuguese. The BP minicorpus is comprised of texts from the informal component of the C-ORAL-BRASIL I corpus (RASO; MELLO, 2012), which together with the other C-ORAL-BRASIL corpora (RASO et al., forthcoming) constitute the Brazilian counterpart of the C-ORAL-ROM.

The IT and BP minicorpora constitute balanced, scaled-down versions of their respective source corpus, with the additional benefit of featuring IS annotation. This kind of annotation, as discussed below, is very laborious and can only be performed manually. This makes it unfeasible to annotate the entire source corpora. The informationally annotated minicorpora have been devised as a strategy to allow IS studies to be conducted within the L-AcT framework, which requires spontaneous speech data to be organized in a specific manner.

[33] For the second version of the IT minicorpus, see Raso and Bossaglia (forthcoming).

Likewise, the AE minicorpus has been created in order to allow the study of IS aspects according to L-AcT, and it is comparable to the IT and BP minicorpora.

Prior to conducting the present study, we carefully revised the AE and BP minicorpora, concentrating on the prosodic and informational annotations. Thus, the AE and BP data that were analyzed have been gathered from the revised versions of the AE and BP minicorpus. The IT data come exclusively from the set of dialogues of the new version of the IT minicorpus. The monologues could not be used because they had not fully prepared in time to be used in this study.

Regarding the EP data, we had to resort to an *ad hoc* solution, since there is no informationally tagged corpus of this variety of Portuguese. Thus we used an unbalanced sample of topic information units (TOP) taken by ROCHA (2012) from the EP component of the C-ORAL-ROM.

This chapter describes the minicorpora as well as other aspects related to the data used in this study, providing information on the context from which the data come, the main features of the minicorpora and the revision process to which they were subjected. The next section outlines the C-ORAL-BRASIL project, which has fostered the conditions for the creation of the BP and the AE minicorpora and for the development of this research.

4.1 The C-ORAL-BRASIL project and its main corpora

The C-ORAL-BRASIL project is dedicated to the study of spontaneous speech on different linguistic levels and to the compilation of speech corpora following the standards established by L-AcT. The project is led by professors T. Raso and H. Mello at the *Laboratory of Empirical and Experimental Linguistic Studies* (LEEL[34]) at the Federal University of Minas Gerais Faculty of Letters. The project is a division of the C-ORAL-ROM, an international consortium for the compilation of speech corpora and for corpus-based studies (CRESTI; MONEGLIA, 2005).

The C-ORAL-BRASIL I and II corpora (RASO; MELLO, 2012, RASO et al., forthcoming) comprise the most relevant accomplishments of the project in terms of speech corpora compilation. These corpora constitute balanced and representative samples of BP spontaneous speech. They focus on the BP variety spoken in the state of Minas Gerais, but they

[34] Portuguese abbreviation for *Laboratório de Estudos Empíricos e Experimentais da Linguagem*.

also document other varieties of BP, particularly in the case of the C-ORAL-BRASIL II that contains texts recorded from television and radio programs.

In total, the C-ORAL-BRASIL I and II have 360 texts and nearly 500,000 words documenting a rich variety of both formal and informal interactions encompassing monologues, dialogues, and conversations that take place in both family/private and public sociological contexts. The C-ORAL-BRASIL corpora feature:

- transcriptions following the CHAT[35] system (MACWHINNEY, 2000) available in TXT and XML formats
- annotation of prosodic boundaries (terminal and non-terminal breaks, retractions, and interruptions)
- annotation of syntactic functions with the PALAVRAS parser (BICK, 2012, 2014)
- utterance-level, text-to-speech alignment carried out with Winpitch (MARTIN, 2005)
- recordings in WAV format exhibiting good to outstanding acoustic quality
- architecture comparable to the C-ORAL-ROM multilanguage corpus, thus enabling crosslinguistic studies.

As mentioned earlier, the theoretical underpinnings underlying the design of the C-ORAL-BRASIL corpora are those established by L-AcT (see chapter 2), a framework for the study of spoken language that envisages the speech act and the organization of information in an utterance as intrinsically related to prosody. In this framework, prosodic parameters are regarded as responsible for signaling illocutionary values as well as marking the internal organization of utterances. These functions are carried out by means of prosodic breaks and a combination of prosodic parameters that vary according to the function considered (see chapter 2). The prosodic annotation of the C-ORAL-BRASIL and C-ORAL-ROM corpora encompass the following:

- terminal prosodic breaks ("//"), which communicate utterance completion
- non-terminal prosodic breaks ("/"), which marks non-autonomous prosodic units inside the same utterance

[35] Acronym that stands for *Codes for the Human Analysis of Transcripts*.

- retractions ("[/n]")[36], which can be thought of as "prosodic erasers" that the speaker can resort to in order to make minor corrections to the locutive content of utterance without having to give up on the utterance
- interruptions ("+"), which constitute prosodic breaks that mark abandoned utterances.

Identifying prosodic units is a preliminary step for carrying out studies on information structure according to L-AcT. Studies of this kind only become possible once the information function of each prosodic unit in a given sample is identified and annotated. But informational annotation constitutes a laborious task that takes a considerable amount of time and the collaboration of trained individuals. This means that corpora like those of the C-ORAL family can hardly receive full informational annotation. Consider, for instance, that informal part of the C-ORAL-BRASIL corpus alone contains nearly 62,000 prosodic units. Therefore, scaled-down versions of these corpora have been created so as to afford the necessary conditions for information structure annotation. These scaled-down versions correspond to the minicorpora referred to in the beginning of this chapter.

The minicorpora of the C-ORAL family maintain the same architecture of the corpora from which they come. The BP minicorpus (RASO; MITTMANN, 2011, RASO; CRESTI, 2012) comes from the C-ORAL-BRASIL I corpus, while both versions of the IT minicorpus (CRESTI; RASO, 2012, RASO; BOSSAGLIA, forthcoming) come from the informal part of the Italian component of the C-ORAL-ROM corpora.

Both the IT and BP minicorpora are composed of an equivalent proportion of monologues, dialogues, and conversations, and they are also somewhat balanced in terms of the distribution of their texts according to sociological context (family/private and public). A minicorpus of Spanish[37] has recently been created (MARTÍNEZ; SOMACARRERA, 2018), but since it has not been used in the present study it will not be described here. All the minicorpora of the C-ORAL family can be freely accessed on the DB-IPIC website [38]. As for the AE minicorpus (CAVALCANTE; RAMOS, 2016), it comparable to the IT and BP minicorpora. Additional information on these three minicorpora is provided in dedicated sections below.

---

[36] The "n." stands for a digit representing the number of retracted words.
[37] The Spanish minicorpus can be accessed at http://www.lablita.it/app/dbipic/.
[38] *Database for Information Patterning Interlinguistic comparison*: http://www.lablita.it/app/dbipic/.

4.2 The IT and BT minicorpora

The IT[39] and the BP minicorpora were conceived to advance the understanding of illocutions and information structure in spontaneous speech (PANUNZI; MITTMANN, 2014) following the approach of L-AcT. Language resources envisaged for such a purpose must capture the widest possible variety of speech acts. As discussed in chapter 2, speech act variation goes hand in hand with the type of activity performed by speakers as they interact and with the setting in which the interaction takes place (MONEGLIA, 2011). In view of that, the texts that make up these two C-ORAL family minicorpora were carefully selected so as to ensure the widest possible variation of situations – this of course also being the case for the C-ORAL family corpora as a whole. And because prosody plays such an important role in speech, the acoustic quality of the recordings selected to make up the minicorpora was also very carefully considered.

The levels of acoustic quality of the BP minicorpus, which is entirely made up of digital recordings, were classified according to the levels shown in Table 4.1.

Table 4.1: Acoustic-quality levels and description

Level	Description
A	Very high quality. Almost no voice overlapping or background noise. Trustable f0 computation for most of the file.
AB	High quality. Low voice overlapping or background noise. Trustable f0 computation for most of the file.
B	Medium quality. Some voice overlapping or background noise. Trustable f0 computation for most of the file.
BC	Mid-low quality. Some voice overlapping or background noise. Trustable f0 computation for at least 60% of the file. Audio is clear for listening throughout the entire file.
C	Low quality. Some voice overlapping or background noise. Trustable F0 computation for at least 60% of the file. Some portions of the audio may not be clear for listening.

Source: Cavalcante and Ramos (2016, p. 114)

The number of texts of the BP minicorpus in each of the acoustic-quality levels shown in Table 1 is as follows:

– Quality level A: 9 texts

[39] This section presents the first version of the IT minicorpus. For a description of its second version, which at the time of writing had not been finished, see Raso and Bossaglia (forthcoming).

- Quality level AB: 9 texts
- Quality level B: 1 text
- Quality level C: 1 text

As can be seen, 90% of the texts in the BP minicorpus are of high to very high quality.

Regarding the IT minicorpus, its three acoustic-quality levels were established based on criteria that are different from those adopted in the C-ORAL-BRASIL project. The IT texts were classified as follows: group A, encompassing both digital and analogue recordings exhibiting high acoustic quality; group B, with digital and analogue recordings of intermediate quality; and group C, with all analogue recordings of low acoustic quality[40]. The number of texts in each of these quality levels are:

- Quality level A: 12 texts
- Quality level B: 7 text
- Quality level C: 1 text

The overall acoustic quality of the BP minicorpus is considerably higher than that of the IT minicorpus, which was one of the main motivations for the second version of the IT minicorpus to be created.

Each minicorpus contains 20 texts totaling about 34,500 words[41], one-third of which corresponding to monologues and two-thirds to dialogues and conversations. The two minicorpora differ slightly in the distribution of texts according to sociological contexts (family/private and public): while the BP minicorpus has 15 family/private texts and 5 public ones, the IT minicorpus has 14 and 6, respectively.

Table 4.2 below provides a summary of the situations documented in the minicorpora. As can be seen, they feature a variety of contexts, which is particularly true of dialogues and conversations. These include situations in where participants perform various types of activities (e.g. shopping, driving, cleaning, game playing etc.). Most monologues, on the other hand, constitute narratives, with the exception of a few interviews. This is because prototypical

[40] For more details about acoustic quality classification of the C-ORAL family corpora, see Cresti and Moneglia (2005) and Raso and Mello (2012).
[41] According to counts reported in Panunzi and Mittmann (2014) for the first version of the IT minicorpus.

monologues, understood as interactions in which the production of content is mostly controlled by one speaker alone, occur only occasionally in spontaneous speech.

Table 4.2: The IT and BP minicorpora communicative situations

IT minicorpus	BP minicorpus
Monologues	
1. Interview with an old partisan at his home	1. Story about a snake
2. Elderly woman tells life story to relatives	2. Family stories by grandmother
3. Story about honeymoon	3. Entertaining stories by a father
4. An after-dinner travel tale to friends	4. A woman's experience in the hospital
5. Interview with a retired traveling salesman	5. Story about daughter adoption
6. Political speech at a political-party meeting	6. A man's career path
7. Explanation about work at an office work	7. Interview with public school teacher
8. Interview with a municipality employee	
Dialogues	
1. Interview with artisan at leather workshop	1. Two friends shop for groceries
2. Friends making a cake	2. Chat while packing equipment
3. Beautician and customer at beauty parlor	3. A couple takes a car trip
4. Developing photos in a darkroom	4. Maids doing the dishes
5. Father teaches daughter how to drive	5. Broker shows apartment to sister
6. On insurance policy	6. Engineer and mason at site
7. Teachers' meeting	7. Shoe store sales encounter
Conversations	
1. Browsing through family photos	1. On a soccer championship
2. Explaining Mastermind	2. On an upcoming marriage
3. Family talks with child while preparing lunch	3. Friends play snooker
4. Meeting of a voluntary association	4. Friends play Pictionary
5. Chat at a hardware store while shopping	5. Job explanation at a blood bank
	6. Political meeting

Adapted from: Panunzi and Mittmann (2014, p. 142-143)

In addition to variation of situations and acoustic quality, the minicorpora also feature many different speakers, equilibrium of male and female voices, age groups, and levels of education. The next section provides details about the AE minicorpus.

4.3 The American English minicorpus

The AE minicorpus, like its IT and BP counterparts, is composed of 20 spontaneous speech recordings with the respective transcripts totaling 30,400 words[42]. As said earlier, its

[42] The word counts of the AE texts reported here differ from those in Cavalcante and Ramos (2016). This is because we have considered contracted forms, e.g. *'s* (is), *'re* (are), *'nt* (not) etc., as individual words, contrary to what had been done for the counts reported in that paper.

texts were taken from the *Santa Barbara Corpus of Spoken American English* (SBCSAE; DU BOIS et al., 2000-2005), which was selected as the source corpus essentially because:

- it is a spontaneous speech corpus, i.e. it features speech whose planning and execution take place simultaneously
- it is a corpus of English, a non-Romance language of wide academic reach
- it is available online licensed under a Creative Commons attribution[43]
- it contains a variety of communicative situations and
- it has recordings of good acoustic quality.

As can been seen, the main factors considered for the selection of texts to compose the AE minicorpus were variation of communicative situations and activities performed while talking, acoustic quality, and, albeit to a lesser extent, equilibrium of male and female voices. Like the other C-ORAL family minicorpora, the AE minicorpus contains monologues, dialogues, and conversations, which are outlined in Table 4.3 below.

As cleared up earlier, a monologue is considered to be an interaction in which one speaker alone controls most of the textual production. Monologue 7 in Table 4.3 below, for example, has 7 participants and 2,853 words. However, 82.7% of the total number of words are uttered by only one speaker. A dialogue, on the other hand, is an interaction in which two speakers share, more or less equally, the textual production, with minor interventions from third parties also being possible. A conversation is like a dialogue, except that the textual production is carried out by at least three speakers.

Much like in the other minicorpora, the words in the AE monologues represent about 35% of the total. The remaining 65% of words are divided between dialogues (40%) and conversations (25%). When it comes to equilibrium of male and female voices, the AE minicorpus is balanced, with 49% of words produced by male and 51% by female speakers. For more details about the AE minicorpus, e.g. participants' age and level of education, see Cavalcante and Ramos (2016).

Table 4.3: American English minicorpus communicative situations

AE minicorpus
Monologues

[43] CC BY-ND 3.0 US.

1. A student explains her studies in equine science on a house trailer
2. Two friends/co-workers talk about their interests at work
3. Two cousins chat at home after a long time apart
4. A man talks about his experiences as a gay man at home
5. Two friends talk as they watch TV at home
6. Two male friends chat about science and human nature at home
7. Meeting about penguins at an aquarium (public interaction)
Dialogues
1. Two cousins chat at home
2. A couple lying in bed talk about a book
3. Mother and daughter at home talk after work
4. A man and a woman talk on a visit to her ranch
5. A couple plays Hearts in a summer house
6. Experienced air-traffic controller and new colleague at work (public interaction)
7. Homeowner and engineer talking about air-conditioning systems (public interaction)
8. Salesman and buyer discussing different types of tape decks (public interaction)
Conversations
1. Three friends chat about traveling, health and vitamins in the living room
2. Two sisters and mother talk at a restaurant table as they decide on what to eat
3. Friends talk at a block party
4. Family members chat at a birthday party
5. Friends talk at a dinner party

Source: Cavalcante and Ramos (2016, p. 112)

Table 4.3 shows the different situations contained in the AE minicorpus, which include parties (conversations 3-5), a professional meeting at an aquarium (monologue 7), a couple playing Hearts (dialogue 5), a married couple lying in bed and talking about a book (dialogue 2), a sales encounter (dialogue 8), and so on. This variation, however, does not always means that speakers are performing different activities besides talking. Exceptions are dialogues 3, 4, 5, 6, 7, 8, conversation 2, and, to a certain extent, monologues 5 and 7. This means that about half of the interactions in the minicorpus is not expected to induce as much illocutionary variation.

The acoustic quality of the recordings in the AE minicorpus was established according to the C-ORAL-BRASIL criteria (see Table 4.1). Out of the 20 recordings of the AE minicorpus, 14 exhibit medium (B) to very high (A) acoustic quality. The remaining 6 recordings are of either mid-low (BC) or low (C) acoustic quality. The number of recordings in each quality level is as follows.

- Quality level A: 1 text
- Quality level AB: 3 texts

- Quality level B: 10 texts
- Quality level BC: 2 texts
- Quality level C: 4 texts

The C-ORAL-BRASIL transcription conventions (see MELLO et al., 2012), some of which are shown in Table 4.4 below, were implemented in the SBCAE texts selected to compose the AE minicorpus. These conventions are theoretically and methodologically motivated, their implementation being part of the measures taken in order to make the AE minicorpus comparable to the other C-ORAL family minicorpora.

Table 4.4: Symbols used in the AE minicorpus transcriptions

Symbol	Meaning
hhh	Paralinguistic noise, e.g. laughs, coughs and throat clearings.
&he	Hesitation or time-taking vocalization
&	Interrupted word; the & sign is put immediately before the interrupted word
< >	Overlapping sequence
yyy	Anonymized information (person, institution, telephone numbers, etc.)
xxx	One incomprehensible word
yyyy	More than one incomprehensible word

Other conventions that were implemented in the AE minicorpus include:

- transcription of alphabet letters pronounced in isolation according to the way they are pronounced (e.g. the letter “a” is transcribed *ey*, “b” is transcribed *bee*, “c” is transcribed *cee* etc.)
- transcription of initialisms as if they were words (e.g. in dialogue 6, VFR (Visual Flight Rules) is pronounced [vi:ɛf’ar] and thus transcribed as *veeefar*)
- transcription of acronyms as a sequence of capital letters (e.g. Metropolitan Museum, usually pronounced [mɛt], is transcribed MET).

The next section deals with details about the prosodic annotation and the text-to-speech alignment of AE minicorpus

4.3.1 Prosodic annotation and text-to-speech alignment of the AE minicorpus

Complying with the methodological standards of L-AcT, the prosodic annotation of the AE minicorpus was carried out on the basis of auditorily salient prosodic cues. The most important aspects that this level of annotation involves is

- detection of terminal prosodic boundaries, which mark the completion of an utterance, i.e. a pragmatically autonomous speech unit
- detection of non-terminal prosodic boundaries, which are responsible for the internal organization of compound utterances, i.e. those that hold more than one prosodic unit.

The tags used for the prosodic annotation correspond to most of the ones used in the C-ORAL family corpora. They have been mentioned in section 4.1 and are repeated in Table 4.5 below for convenience.

Table 4.5: Symbols used for the prosodic annotation of the AE minicorpus

Symbol	Value
//	terminal prosodic break
/	non-terminal prosodic break
[/n.]	retraction
+	interrupted utterance
<	beginning of overlapping sequence
>	end of overlapping sequence

Besides identifying terminal and non-terminal breaks, the prosodic annotation also identifies retraction and interruption phenomena. A *retraction* "[/n.]" corresponds to false starts, hesitations, performance slips and attendant prosodic discontinuities that act upon a prosodic unit without resulting in the total invalidation of the ongoing utterance. An *interruption* "+", on the other hand, occur when the speaker abandons the utterance altogether, in which case she may either yield the turn to another speaker or simply start another utterance. Finally, overlapping speech is enclosed in angle brackets "< >".

The prosodic annotation is a precondition for text-to-speech alignment. All the C-ORAL family corpora are aligned with the Winpitch software (MARTIN, 2005), taking the utterance as the unit for alignment. The output of the alignment process carried out with Winpitch includes an XML file which, together with the respective audio file, allows one to easily navigate the transcription and sound of the aligned text.

The AE minicorpus, just like its IT and BP counterparts, are full aligned with Winpitch, which makes them particularly suitable for the type of analyses carried out in this study, since

all of them require an easy access to both text and sound sources. Figure 4.1 below shows a screenshot of an AE minicorpus text (XML and WAV files) in a Winpitch Text window.

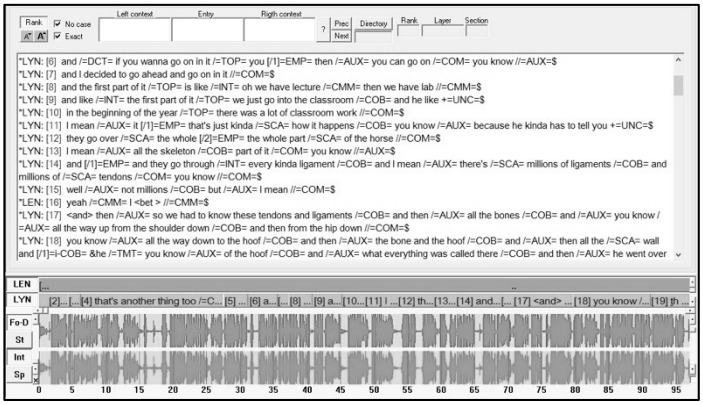


Figure 4.1 – Screenshot of Winpitch Text window with wave forms in the bottom

All the aligned utterances can be easily heard by simply clicking on them in the window shown in Figure 4.1.The alignment software also allows the visualization of acoustic parameters my means of spectrograms, waveforms, f0 and intensity curves, and so forth.

The following section deals with issues related to the informational annotation of the AE minicorpus.

4.3.2 Information-structure annotation and revision

The AE minicorpus, like the other minicorpora of the C-ORAL family, features information-structure annotation, which was carried out according to the *Information Patterning Hypothesis* (CRESTI; MONEGLIA, 2010). This level of annotation involves the manual assignment of tags indicating the informational function of each prosodic unit in the minicorpus, thus requiring a considerable amount of time. Table 4.6 below shows the tags used in the informational annotation of the C-ORAL family minicorpora. The information units corresponding to these tags are discussed and exemplified in chapter 2.

The annotation of informational functions involves an analysis of the prosodic features of each information unit both auditorily and visually, so that the informational function be accurately identified, and the tag properly assigned. As seen in Table 4.5, the information tags consist of a sequence of three capital letters enclosed between two equal signs and placed right after the sign indicating the prosodic break. Additional characters are used for annotating units in reported speech, whose three-letter tags are accompanied by an underscore character

followed by a lowercase *r* (e.g. =XXX_r=). In addition, tags indicating interrupted units are preceded by the sign "i-", as shown in (4.1) below, which features an annotated excerpt from the AE minicorpus containing, among other elements, the additional characters that have just been mentioned.

(4.1) afamcv03_283 (ex4.1.wav):
so /=AUX= I went to this [/1]=i-COB= &m [/1]=EMP= what I thought was my friend /=PAR= &he /=TMT= &th [/1]=EMP= this navy captain down at /=SCA= &he /=TMT= naval headquarters /=COB= I said /=INT= this is terribly awkward /=COB_r= I've just been promoted from third mate /=CMM_r= to second mate /=CMM_r= and [/1]=EMP= and could we possibly postpone these orders for a little bit //=COM_r=

Dialogic units in the AE minicorpus have been, for the most part, tagged AUX. This is because in order to be further distinguished, dialogic units – so far, six different classes of such units have been identified, as can be seen in chapter 2 and Table 4.6 – require a level of acoustic quality the SBCSAE recordings often fail to have due to the technology limitations at the time the corpus was compiled.

Since the prosodic and the informational annotation are carried out manually, they are susceptible to errors and thus require thorough and time-consuming revisions. Both the AE and BP minicorpus have been revised for the studies conducted for this dissertation. As for the IT minicorpus, instead of submitting it to a revision, it has been decided that it would be more suitable to create a whole new version of it, which, as already mentioned, is underway (RASO; BOSSAGLIA, forthcoming).

The revision of the prosodic and informational annotation of the AE and BP minicorpus was done simultaneously by three annotators, namely this author, his dissertation adviser, and Prof. Giulia Bossaglia, who is another member of the LEEL team. Winpitch and Praat (BOERSMA; WEENINK, 2015) were used in the process. The revisers went over every prosodic unit in the two minicorpora checking for annotation inconsistencies and mistakes relative to the nature and placement of prosodic breaks as well as errors related to the assignment of informational units.

Table 4.6: Tagset used for the annotation of informational functions

Name	Tag	Description
		Illocutionary textual units
Comment	COM	illocutionary unit; necessary and sufficient for the performance of an utterance.
Multiple Comment	CMM	an illocution in a patterned chain consisting of at least another CMM and performing one conventional rhetoric effect.
Bound Comment	COB	an illocution featuring a prosodic signal of continuity .
		Non-illocutionary textual units
Topic	TOP	unit that supplies a cognitive domain for the illocution, allowing the utterance to be displaced from its immediate context.
Topic List	TPL	unit in a semantically connected sequence that supply one single domain of identification for the illocution.
Appendix of COM	APC	unit that supplies additional text for an illocutionary unit.
Appendix of TOP	APT	unit that supplies additional text for a TOP unit.
Parenthetic	PAR	unit with metalinguistic and/or modal content that aids in the interpretation of the utterance or part thereof.
Parenthetic List	PRL	unit in a semantically and syntactically connected sequence supplying metalinguistic and modal content for the interpretation of the utterance or part thereof.
Locutive Introducer	INT	unit marking a shift of pragmatic coordinates.
		Dialogic units
Incipit	INP	unit that begins dialogic turns and utterances and that imparts some sort of affective contrast.
Conative	CNT	unit that that pushes the interlocutor to do or stop doing something.
Phatic	PHA	unit that manages the communicative channel ensuring its maintenance.
Allocutive	ALL	unit that specifies the addressee enacting social cohesion.
Expressive	EXP	unit that supplies support by marking social affiliation.
Discourse Connector	DCT	unit that connects different parts of the discourse, indicating its continuation.
		Other units
Scanning Unit	SCA	prosodic unit with no signal of information function holding part of the syntactic-semantic content of a textual information unit.
Interrupted unit	i-TAG	momentarily suspended textual unit eventually picked up after the realization of another information unit.
Reported unit	TAG_r	unit in reported speech
Empty unit	EMP	fully retracted prosodic unit
Time Taking	TMT	unit consisting of a filled pause
Unclassified	UNC	unit whose function couldn't be determined

In the following section, more details about the makeup of the BP and AE minicorpora are presented.

4.3.3 Comparing the BP and AE minicorpora

As already said, the two minicorpus have undergone a careful revision, concentrating on the prosodic and informational annotations. The AE minicorpus also received a few minor transcription adjustments on contracted forms so as to facilitate word counts. The tallies

presented here differ slightly from other sources (e.g. PANUNZI; MITTMANN, 2014 and CAVALCANTE; RAMOS, 2016). This is primarily due to the revision, which yielded differences in the number of utterances and prosodic units. Particularly in the case of the AE minicorpus, the revision also produced differences in the number of words, as contracted forms were counted as individual words.

Table 4.7: Number of words, texts, and utterances in the BP and AE minicorpora

		BP minicorpus			**AE minicorpus**		
Context	Type	Texts	Words	Utterances*	Texts	Words	Utterances*
family/private	mn	6	8,672	759	6	7,752	637
	dl	5	8,372	1,706	5	7,614	883
	cv	4	6,393	1,201	5	7,500	1,083
public	mn	1	1,637	126	1	2,853	322
	dl	2	3,044	540	3	4,742	556
	cv	2	3,459	536	0	0	0
Total		20	31,577	4,868	20	30,461	3,481

* Interrupted utterances are not included in these counts.

The two minicorpora are very similar in terms of number of words: 31,577 in the BP minicorpus and 30,461 in the AE minicorpus (see Table 4.7). Although the AE minicorpus has no public conversation, its public-context part, containing 7,596 words, is similar to the equivalent part of the BP minicorpus, which contains 8,140 words. Moreover, both minicorpora are balanced in terms of word counts in monologues, which hold about one third of the words of each minicorpus, and dialogues and conversations considered together, which hold about two thirds of the words of each minicorpus.

This important because the structural differences between monologues and multiple-party interactions (i.e. dialogues and conversations) are much more relevant than the differences between dialogues and conversations. Many of the differences between monologues and multiple-party interactions are due to the fact that monologues have a lower degree of interactivity in the sense that the speaker in a monologue is for the most part involved with nothing but the act of speaking itself, whereas the speakers involved in a multiple-party interaction are more likely to be also involved in non-verbal activities (see CRESTI, 2005, PANNUNZI; MITTMANN, 2014). Thus, the monologues hold 33% of the words in the BP minicorpus and 35% in the AE minicorpus, whereas multiple-party interactions hold 67% (BP) and 65% (AE).

Table 4.7 also shows tallies relative to concluded utterances. As can be seen, most utterances are found in texts featuring multiple-party interactions, which constitutes a desirable circumstance, given that many interesting pragmatic phenomena – variation of illocutionary types in particular – are expected to be found in dialogues and conversations. The distribution of concluded utterances in the BP minicorpus is 18% in monologues and 82% in dialogues and conversations. In the AE minicorpus, the distribution is 28% of concluded utterances in monologues and 72% in dialogues and conversations.

This difference between the minicorpora is partially accounted for by the fact that many of the dialogic texts in the AE minicorpus correspond to interactions in which speakers are exclusively engaged the verbal activity. This results in texts that are more similar to monologues, in the sense that such dialogues tend to exhibit relatively longer stretches of speech by one individual alone and lower interactivity. Consequently, the utterances in dialogic texts of this type tend be fewer in number, while being likely to hold a greater number of words and prosodic/information units.

Table 4.8 shows the distribution in the minicorpora of words per utterance, per information unit, per dialogic turn, and per minute. Considering the totals, the ratios in the AE minicorpus are consistently greater than in the BP minicorpus. This is likely, at least in part, to be due to the rhythmic structure of the of the English language, which favors syllable reduction, and the average number of syllables per words in English, which tend to be small. But when the ratios relative to dialogues and conversations and the ratios relative to monologues are considered separately, the effect of interactivity seems to become more apparent. Thus, we see that dialogues and conversations in the AE minicorpus, because of their lower degree of interactivity, have ratios that are considerably greater than dialogues and conversations in the BP minicorpus. At the same time, BP monologues exhibit greater ratios than AE monologues, which suggests that the monologues in the BP minicorpus are more prototypical – i.e. lower in interactivity – than those in the AE minicorpus.

Table 4.8: Number of words per IU, utterance, turn and minute in the minicorpora

BP minicorpus						
Section	Words	Words/IU	Words/utt	Words/Turn	Words/min	Dur. (min)
mn	10,309	3.1	10.9	41.2	185	55.7
dl+cv	21,268	2.8	4.9	8.0	138	154.0
Total	31,577	2.9	5.9	10.8	151	209.8
AE minicorpus						
AE mini	Words	Words/IU	Words/utt	Words/Turn	Words/min	Dur. (min)

mn	10,605	3.4	9.6	24.7	194	54.6
dl+cv	19,856	3.3	6.9	11.0	215	92.4
Total	30,461	3.3	7.6	13.7	207	147.0

The section below provides information regarding the distribution of information units in the minicorpora.

4.3.4 Information units: the general picture in the AE and BP minicorpora

Table 4.9 below shows the frequency of the different types of information units in the two minicorpora. The illocutionary group encompasses the three types of textual information unit that carry illocution, namely COM, COB, and CMM, and represent 60% of the information units in the BP minicorpus and 54.2% in the AE minicorpus. As can be seen in Table 4.9, monologues tend to have smaller proportions of illocutionary units as compared to dialogues and conversations. This is because monologues tend to be more complex from the textual point of view, exhibiting a higher proportion of units that expand the semantic content of utterances.

The non-illocutionary textual group, which corresponds to the lines named “Textual” in Table 4.9, encompasses TOP, PAR, INT, APC, and APT. They represent 11% and 9.7 percent of information units in the BP and AE minicorpora, respectively. The dialogic group, on the other hand, holds the units that are equivalent to discourse markers in other approaches, encompassing PHA, CNT, INT, ALL, EXP, and DCT, and have virtually the same proportion in the two minicorpora (about 11%).

SCAs constitute prosodic units that carry locutive content of textual units and have no informational function on their own. They represent 8% of information units in the BP minicorpus and 9.2% in the AE minicorpus. TMTs constitute prosodic units used as filled pauses (1% in the BP minicorpus and 4.1% in the AE minicorpus), while EMPs correspond to those units whose words are retracted (6% in the BP minicorpus and 7.1% in the AE minicorpus). Finally, the UNC group correspond to the prosodic units which could not be reliably classified.

As mentioned earlier, the two minicorpora differ in terms of the overall proportion of illocutionary units, as these represent 60% of the units in the BP minicorpus and 54.2% in the AE minicorpus. However, the proportion of illocutionary units in monologues is practically the same in the two minicorpus (about 50%). The same cannot be said of the dialogic sections (i.e. dl+cv), as the illocutionary group represent 65% of information units in the BP dialogic section

but 56% in the AE counterpart. As mentioned above, AE dialogues exhibit a lower degree of interactivity, which in part explains why the dialogic section of the AE minicorpus has a proportion of illocutionary units closer to that of its monologic section.

The slightly higher proportion of UNCs in the AE minicorpus is likely the result of the lower quality of the minicorpus' acoustic source and the higher incidence of speech overlapping than in the BP minicorpus. In addition, the informational annotation of the AE minicorpus was carried out by non-native speakers of English, which has certainly made the classification of some prosodic units harder. As for TMTs, their higher proportion in the AE minicorpus might be an indication that filled pauses have a special role in AE as a discourse strategy, which may constitute an interesting topic for investigation.

Table 4.9: Frequency of information unit (IU) groups in the minicorpora

BP minicorpus						
IU groups	monologues		dl + cv		Total	
Illocutionary	1,636	50%	4,937	65%	6,573	60%
Textual	520	16%	663	9%	1,183	11%
Dialogic	334	10%	820	11%	1,154	11%
SCA	422	13%	474	6%	896	8%
TMT	71	2%	73	1%	144	1%
EMP	219	7%	413	5%	632	6%
UNC	51	2%	262	3%	313	3%
Total	3,253	100%	7,642	100%	10,895	100%
AE minicorpus						
IU groups	monologues		dl + cv		Total	
Illocutionary	1,608	50.6%	3,375	56.0%	4,983	54.2%
Textual	410	12.9%	485	8.0%	895	9.7%
Dialogic	387	12.2%	653	10.8%	1,040	11.3%
SCA	323	10.2%	527	8.7%	850	9.2%
TMT	120	3.8%	255	4.2%	375	4.1%
EMP	206	6.5%	449	7.5%	655	7.1%
UNC	121	3.8%	281	4.7%	402	4.4%
Total	3,175	100.0%	6,025	100.0%	9,200	100.0%

As shown in chapter 2, there are different types of utterances. The simplest type contains only one prosodic unit – invariably COM – dedicated to carrying the illocution. The frequency of simple utterances correlates with the type of interaction, with monologues tending to have fewer simple utterances than dialogic texts. As Table 4.10 below shows, simple utterances represent 60.5% of BP utterances in dialogic interactions and 59.2% of the utterances in corresponding part of the AE minicorpus. But when monologues are considered, the incidence

of simple utterances shows a greater difference: 37.4% in the BP monologues and 45.2% in the AE monologues.

Table 4.10: Frequency of different types of utterances in the minicorpora

BP minicorpus									
Section	simple		aux+ill		text+(aux)+ill		ill+ill		Total
mn	356	37.4%	59	6.2%	445	46.7%	92	9.7%	952
cv+dl	2,655	60.5%	150	3.4%	795	18.1%	789	18.0%	4,389
Total	3,011	56.4%	209	3.9%	1,240	23.2%	881	16.5%	5,341
AE minicorpus									
AE mini	simple		aux+ill		text+(aux)+ill		ill+ill		Total
mn	499	45.2%	74	6.7%	404	36.6%	127	11.5%	1,104
cv+dl	1,703	59.2%	149	5.2%	662	23.0%	364	12.6%	2,878
Total	2,202	55.3%	223	5.6%	1,066	26.8%	491	12.3%	3,982

Compound utterances, i.e. utterances that have more than one prosodic unit, can exhibit different types of arrangements. First of all, there are compound utterances whose COM unit is accompanied by one or more dialogic units. These are shown under the aux+ill headings in Table 4.10. As can be seen, in both the BP and AE minicorpora, the frequency aux+ill utterances is higher in monologues (6.2% and 6.7%, respectively) than in the multiple-party interactions (3.4% and 5.2%, respectively), though the difference in frequency is smaller in AE than in BP.

Another way in which utterances can be complex is by having a COM unit accompanied by at least one non-illocutionary textual unit, with the optional presence dialogic information units. The frequencies of this type of compound utterance are shown in the column labeled text+(aux)+ill in Table 4.10, which shows that this type of utterance makes up 46.7% of BP monologues and 36.6% of AE monologues. As mentioned earlier in this chapter (see chapter 2 for details), the non-illocutionary textual units are TOP, PAR, INT, APT, and APC. These units compose the syntactic-semantic content of the utterance, but do not convey any illocutionary value.

In the AE dialogic texts, the frequency of utterances with one COM unit plus at least one non-illocutionary textual unit is a little higher than in corresponding texts of the BP minicorpus (23% and 18.1%, respectively). This may be explained by the fact that the dialogic interactions in the AE resource, as said above, contain passages that are less interactive.

Table 4.10 also provides information regarding a type of compound utterance that is made up of more than one illocutionary unit featuring a conventionalized melodic contour that

conveys a rhetorical effect. These are referred to as *compositional illocutionary patterns*, as shown in chapter 2. It must be clarified that utterances counted in the ill+ill column in Table 4.10 correspond only to those compositional illocutionary patterns that are made up exclusively of illocutionary units. Compositional illocutionary patterns containing other information units are counted under the text+(aux)+ill heading. As the table shows, in BP the incidence of compositional illocutionary patterns exclusively made up of illocutionary units is much higher in dialogic texts than in monologues (18% and 9.7% respectively). In AE, on the other hand, this type of utterance exhibits very similar proportions in both dialogues (12.6%) and monologues (11.5%). As other aspects related to the AE minicorpus, this may also be accounted for by the lesser interactivity present in the dialogic part of the AE minicorpus as compared to the corresponding part of the BP minicorpus.

Finally, there are utterances, referred to as stanzas, containing at least two illocutionary units which, unlike the illocutions in a compositional illocutionary pattern, convey no rhetorical effect but are connected to each other by a prosodic signal of continuity. As shown in Table 4.11, stanzas are more frequent in multiple-party interactions. However, the table also indicates that there is a tendency for stanzas in monologues to hold more bound comments (COB) than stanzas in multiple-party interactions.

The fact that AE dialogues and conversations are closer to monologues in terms of the average number of COBs in stanzas may also be accounted for by the lower degree of interactivity in the multiple-party interactions of the AE minicorpus. After all, less interactivity implies more room for textual elaboration as well as for a piecemeal construction of speech content, which may result in a higher incidence of stanzas featuring a greater number of COBs.

Table 4.11: Frequency of stanzas and bound comments in the minicorpora

BP minicorpus					
Section	Number of stanzas		Number of COBs		COB per stanza
mn	265	46.9%	587	57.4%	2.22
dl+cv	300	53.1%	435	42.6%	1.45
Total	565	100.0%	1,022	100.0%	1.81
AE minicorpus					
Section	Number of Stanzas		Number of COBs		COB per Stanza
mn	233	39.0%	468	45.0%	2.01
dl+cv	364	61.0%	573	55.0%	1.57
Total	597	100.0%	1,041	100.0%	1.74

Table 4.12 below shows the frequency of individual types of informational units in the monologues and dialogic texts of the minicorpora. It provides a general picture as to how each type of illocutionary and non-illocutionary textual units are distributed across the minicorpora. Dialogic unit types are not distinguished, since the AE minicorpus, as said earlier, does not allow detailed evaluation of such units due to the overall level of quality of its acoustic sources.

Table 4.12: Frequency of information units (IUs) in the minicorpora. The proportions are give row wise, thus representing the percentage of the total of IUs in monolouges and in multiple-party interactions (dl+cv)

BP minicorpus						
IUs	monologues		dl+cv		Total	
APC	39	34,5%	74	65,5%	113	1,0%
APT	16	66,7%	8	33,3%	24	0,2%
AUX*	334	28,9%	820	71,1%	1,154	10,6%
CMM	264	21,5%	965	78,5%	1,229	11,3%
COB	568	57,2%	425	42,8%	993	9,1%
COM	804	18,5%	3547	81,5%	4,351	39,9%
EMP	219	34,7%	413	65,3%	632	5,8%
INT	118	53,6%	102	46,4%	220	2,0%
PAR	99	50,8%	96	49,2%	195	1,8%
SCA	422	47,1%	474	52,9%	896	8,2%
TMT	71	49,3%	73	50,7%	144	1,3%
TOP	237	39,0%	371	61,0%	608	5,6%
TPL	11	47,8%	12	52,2%	23	0,2%
UNC	51	16,3%	262	83,7%	313	2,9%
Total	3,253	29,9%	7,642	70,1%	10,895	100,0%
AE minicorpus						
IUs	monologues		dl+cv		Total	
APC	13	35,1%	24	64,9%	37	0,4%
APT	10	33,3%	20	66,7%	30	0,3%
AUX*	387	37,2%	653	62,8%	1,040	11,3%
CMM	264	33,7%	520	66,3%	784	8,5%
COB	468	45,0%	572	55,0%	1,040	11,3%
COM	876	27,7%	2,283	72,3%	3,159	34,3%
EMP	206	31,5%	449	68,5%	655	7,1%
INT	105	48,4%	112	51,6%	217	2,4%
PAR	68	47,2%	76	52,8%	144	1,6%
SCA	323	38,0%	527	62,0%	850	9,2%
TMT	120	32,0%	255	68,0%	375	4,1%
TOP	204	45,7%	242	54,3%	446	4,8%
TPL	10	47,6%	11	52,4%	21	0,2%
UNC	121	30,1%	281	69,9%	402	4,4%
Total	3,175	34,5%	6,025	65,5%	9,200	100,0%

* Dialogic information units.

In the next section, the dataset containing the TOP units in Italian and European Portuguese is outlined.

4.4 The Italian and European Portuguese TOP units

The European Portuguese data used in this study come from a set of five texts of approximately 1,500 words from the EP component of the C-ORAL-ROM (NASCIMENTO et al., 2005). This dataset was prepared for one of the first studies on TOP carried out by a member of the LEEL lab (ROCHA, 2012).

The EP dataset consists of four informal texts – one monologue, one conversation, and two dialogues –, and one formal monologue. According to Rocha (2012, p. 107), the selection of texts was based on acoustic quality and the variation of communicative contexts. The texts were screened for occurrences of TOP units, and other information units were not annotated. A sum total of 73 TOPs was found (ROCHA, 2012, p. 135). The TOP units in EP analyzed in this dissertation (see chapters 5 to 7) were taken out of this set of 73 units. Our selection was based on the quality of the acoustic sources. Their prosodic forms underwent a reassessment taking into consideration insights from more recent studies (CAVALCANTE, 2015; RASO et al. 2017).

As for the IT data, it was gathered from the aforementioned new version of the IT minicorpus (RASO; BOSSAGLIA, forthcoming). This version is structurally similar to the first one. The major difference between them is that some texts are being replaced with others featuring better acoustic quality. In addition, the informational and prosodic annotations of the texts that will not be replaced are being carefully revised. At the time the analyses reported in this study, only the dialogues and conversations were ready to be used.

4.5 Final words

This chapter dealt with the corpora used in this study. The next chapter presents an experiment conducted to assess the agreement level among four annotators in a task involving the annotation of TOP units based on spontaneous speech data from C-ORAL-BRASIL II.

5 A test of agreement on TOP detection in naturally occurring data

According to the Language into Act Theory (L-AcT), the topic information unit (TOP) has the function of supplying a cognitive domain for the interpretation of the illocutionary force. It is realized by prosodic units with specific melodic and durational patterns, always occurring to the left of the illocutionary unit. As can be seen, TOP has a semantic-pragmatic facet that corresponds to specific formal aspects.

In order to assess the reliability of using the criteria set forth by L-AcT as guidelines for the detection of TOP in spontaneous speech data, we devised an interrater agreement test using the Kappa statistic (COHEN, 1960; FLEISS, 1971). Our aim was to measure the degree of convergence achieved by annotators using the definition proposed by L-AcT to identify TOP units in naturally occurring data. We began by conducting a pilot test, which is presented in the following section.

5.1 Pilot test

The sample used in the pilot test consists of 50 compound, non-interrupted utterances from the C-ORAL-BRASIL II corpus (RASO; MELLO; FERRARI, forthcoming). By *non-interrupted utterances*, we mean utterances that fully accomplished and therefore marked by a terminal prosodic break (see chapter 2). The utterances were taken from the *Natural Context* section of the corpus, which contains formal face-to-face interactions encompassing lectures, political speeches, court hearings, and so forth, just as the homonymous section of the C-ORAL-ROM (CRESTI; MONEGLIA, 2005), the European counterpart of the C-ORAL-BRASIL. Once selected, the utterances were organized in a text file and their corresponding acoustic sources were saved in separate WAV files.

Five participants, including this author, were chosen among the researchers from LEEL who not only are familiar with working within L-AcT but also have extensive experience with information structure annotation. A protocol was set up for the test, where it was specified that participants did the task in three phases. First, they were supposed to identify all the illocutionary units in the sample, then they had to go over every utterance searching for TOP units. In the third and final step, they should annotate all the remaining prosodic units with a generic tag that did not distinguish their functions.

We used Praat (BOERSMA; WEENICK, 2019), an open source software that, among many other things, allows users to play acoustic files and to visualize prosodic parameters. In

addition to annotating the prosodic units, the participants were asked to register their observations in a separate text file, in case they found any problem with the utterances' transcriptions and prosodic segmentations, which they were not allowed to modify, or with the test in general.

The 50 utterances of the sample added up to 260 prosodic units. The number of TOPs identified by the annotators ranged from 17 to 35 (M = 23.6, SD = 8.0), i.e. about 9.1% of the prosodic units in the sample were on average tagged TOPs, which is consistent with prevalence estimates reported in previous studies (MITTMANN, 2012; CAVALCANTE, 2015).

To measure the strength of the agreement among the annotators, we used the kappa statistic (COHEN, 1960), particularly Fleiss's (1971) kappa, which measures the convergence among any number of raters, unlike Cohen's kappa, which can only be used with two raters. One of the advantages of using the kappa statistic is that it accounts for chance agreement, i.e. it takes into consideration the fact that raters may agree simply by chance. In order to compute the kappa coefficients, we used R (R Core Team) and the *kappam.fleiss* function from the *irr Package* (GAMER; LEMON, 2019).

The kappa coefficient expressing the overall agreement in the pilot test was k = 0.51, which means that if one of the raters has annotated a prosodic unit TOP, a second one has a 51% chance of agreeing with her. According to the benchmarks proposed in Landis and Koch (1977, p. 165), the coefficient k = 0.51 expresses moderate agreement.

Table 5.1: Kappa coefficients *k* ranges and level of agreement

k	**Strength of agreement**
< 0	Poor
0.0 – 0.20	Slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.81 – 1.00	Almost perfect

Source: Landis and Koch (1977, p. 165)

We also computed *k* considering four annotators at a time so as to check whether any of them were influencing the strength of the agreement in a particular way (see Table 5.2 below). We found that, by excluding the annotator identified as E in the table below, the agreement strength increased (k = 0.61, i.e. a substantial agreement). It turned out that the annotator in question had spent a very long time – roughly five years – dealing with information structure

annotation only occasionally, and therefore being unable to keep up with all the advances in the area. Thus, we decided that this person, instead of taking part in the actual test as an annotator, would revise the prosodic segmentation of the utterances to be used in it instead. No similarly expressive change in *k* was observed upon removing any of the other 4 annotators.

Table 5.2: *k* coefficients computed taking four annotators at a time

Annotators	**k**
B, C, D, E	0.45
A, C, D, E	0.46
A, B, D, E	0.49
A, B, C, E	0.50
A, B, C, D	0.61

Full disagreements (i.e. only one annotator classified a given unit as TOP) occurred in 22 cases, full agreements (i.e. all annotators classified the unit as TOP) in 6 cases, and partial agreements (i.e. at least two annotators identified the unit as TOP) in 23 cases. Thus, considering both full and partial agreements together, there were 29 cases in which some level of convergence was observed.

The analysis of the results, which were thoroughly discussed during meetings in which four annotators participated, suggests that some of the disagreements observed were caused by inattention on the part of the annotator regarding the semantic aspect of the definition of TOP. This was observed in cases where a unit with prosodic cues resembling those associated with TOP (see chapters 2 and 6 for details) featured locutive content (i.e. morphosyntactic content) that was not supplying a cognitive domain for the interpretation of the illocution.

The Portuguese expression *então* ‘then/so’ in the utterance below is a case in point. While the melodic curve with which it is realized suggests it constitutes a TOP (see Figure 5.1), from the semantic point of view – and when considered out of context – *então* may either convey a temporal meaning, thus being referential, or fulfill a discourse/dialogic role.

(5.1) bnatte05_48 (ex5.1.wav):
então / o que que a gente vai fazer aqui nesses dois dias //
so / what is it that we are we going to do here during these two days //

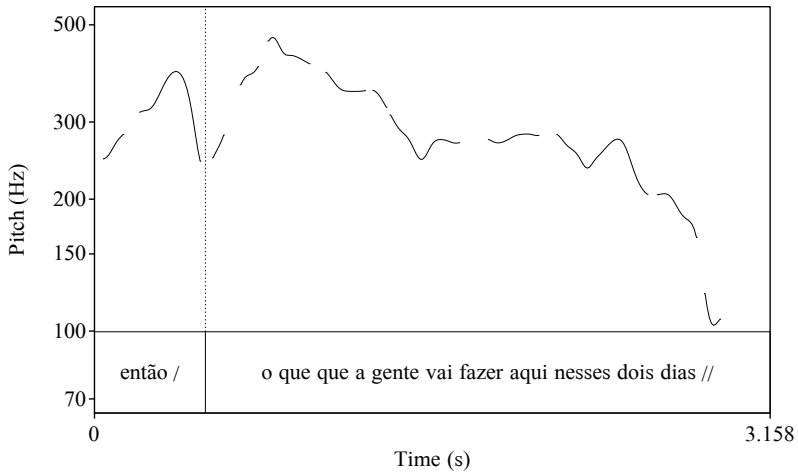


Figure 5.1 – f0 curve of the utterance in ex. (5.1).

In fact, the expression *então* in the utterance above has no referential function, and it constitutes the locutive content to a *dialogic unit*, a group of information units whose function is related to the regulation of the interaction rather than the building of syntactic and semantic content (see RASO, 2014). These units are roughly equivalent to what is referred to as *discourse markers* within other frameworks. As already mentioned, in order for a unit to be TOP, its locutive content must be referential. Thus, *então* would have to mean something along the lines of "after that" in order to qualify as TOP. This requirement is not met, as the example shows, and the English translation clearly suggests.

The analysis of the pilot test results also pointed to another source of inconsistency, namely, the case of units consisting of a reduced number of syllables. Units such as those can truly be TOPs; however, they are likely to have other functions as well. Provided that the unit's locutive content satisfies the referentiality condition, the decision must be made by paying special attention to prosody. For instance, *eles* 'they-masculine' in the example below taken from the pilot test was tagged TOP by one of the annotators. But by listening to the sequence of units in the example with close attention, it becomes clear that *eles* is not functioning as a TOP, since no prosodic nucleus (see chapters 2, 6, and 7) can be identified. The case in question turns out to be a scanning unit (SCA), i.e. a prosodic unit that has no information function in its own right, as it is only a locutive content conveyor, thus being syntactically compositional with the subsequent unit.

(5.2) benatpd09_165 (ex5.2.wav)
eles / não só propõem essa mudança / (...)

they / not only propose this change / (...)

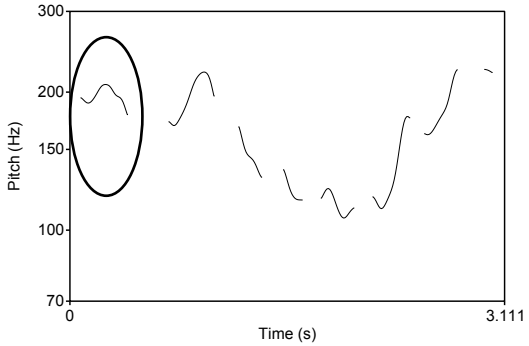


Figure 5.2 – F0 curve of the utterance in ex. (5.2)

In addition, we decided that the participants in the actual test would not be allowed to have access to visual representations of prosodic cues, given that in natural conditions, of course, there is no such possibility. Based on the discussions that we had during the meetings, we set up a protocol (see Appendix A), which will be presented below.

5.2 The kappa test

In this section we discuss the details concerning the agreement test that was designed to assess the reliability of the criteria established by L-AcT for the identification of TOP units in naturally occurring speech data.

5.2.1 Materials and methods

We decided to use Praat in the test, instructing the annotators to disable the Pitch visualization function in the software's Editor window. The annotation steps were the same as in the pilot test, which means that the annotators should first identify the illocutionary units, then TOPs, and finally the remaining units without specifying their information function. Four different tags were specified in the protocol (see Appendix A):

- COM for illocutionary units in regular utterances
- COB for illocutionary units in stanzas (see chapter 2)
- TOP
- and NTP for all other units.

The protocol also contained guidelines for the identification of TOP units, so as to ensure that participants would be using the same criteria. The guidelines contained the following information:

i. TOP is always referential, as it constitutes a cognitive domain for the illocution
ii. Units whose locutive content corresponds to a pronoun alone or that exhibits weak phonetic realization should be very carefully considered: under no circumstance should they be labeled TOP in the absence of a clear prosodic prominence
iii. Prosody is an important aspect, but the semantic criterium, which establishes the referential nature of TOP, must never be forgotten; in other words, to be TOP, a unit has to pass both the semantic and the prosodic tests
iv. As a corollary of (iii), short expressions such as *aí* 'so/then' and *então* 'then/after that' must be clearly anaphoric or have temporal meaning – thus supplying a domain and not functioning as a discourse marker – in order to be TOP
v. Whenever a unit following a TOP shows a clear prosodic contrast with its surroundings, it is likely an information unit other than a TOP (most likely a PAR unit), even if it exhibits a rising movement on its final syllables.

Items (iv) and (v) will be further discussed at the end of this chapter.

For the test, we decided to sample utterances from all the three sections of the C-ORAL-BRASILL II, namely:

- Natural context, a section consisting of face-to-face interactions recorded directly in the context where the exchanges took place
- Media, which contains interactions recorded from TV and radio programs
- Telephonic, which contains short telephone conversations pertaining to both the private and public domains.

We decided to use 100 compound utterances in the test. Concentrating about 15% of the compound utterances of the corpus, the Telephonic section is underrepresented in relation to the Media and Natural context sections, which contain 48% and 37% of the compound utterances of the corpus, respectively. Thus, we decided that 20% of the utterances of our

sample would come from the Telephonic section and the remaining 80% of the utterances would consist of equal portions of Media and Natural Context.

The process of obtaining the utterances included the following steps. Firstly, we excluded those texts which contained less than five utterances. This was particularly important for sampling texts from the Telephonic section, which has some rather short interactions. The reason for excluding such short texts is that we had decided that it was important to provide each *target utterance*, i.e. the one that should receive annotation, along with four surrounding utterances, two immediately before and two immediately after each target utterance. Having excluded the short texts, we randomly selected 40 Media texts, 40 Natural context texts, and 20 Telephonic texts. The 100 target utterances were sampled from these texts, each of which providing only one target utterance.

The utterances were randomly selected exclusively from the pool of non-interrupted, compound utterances in those texts. The selection of texts and utterances was done using R (R Core Team, 2019), and the script prepared for this task can be seen in Appendix B. The utterances were carefully revised to ensure that segmentation and transcription faults were reduced to a minimum. The revision was carried out according to the protocol shown in Appendix C. After the revision, they added up to 541 prosodic units.

Once the utterances had been selected and properly organized in spreadsheets, we proceeded to obtain the audio files. In this phase, we used the *C-ORAL- Search*,[44] a very useful Praat script which allows the automatic retrieval of utterances' acoustic sources in the C-ORAL corpora.

The four annotators, among whom this author is included, received the utterances organized in an XLSX file containing three spreadsheets, one for each section. The target utterances were highlighted, as shown in Table 5.3 below, so that they could be easily distinguished from their surroundings. The audio excerpt of each utterance was saved in a separate WAV file and named as shown in the second column in Table 5.3.

Table 5.3: Example of utterances and audiofiles as annotators received

Utterances	**Audio files**
[14] hhh eu nem sei onde que é na [/1] no [/1] nos States aí //	btelpv30_14.wav
[15] onde que é //	btelpv30_15.wav

[44] The C-ORAL-Search is available at https://vieiramarcelo.wordpress.com/praat-scripts/c-oral-search/.

[16] no / Texas //	btelpv30_16target.wav
[17] ah / tá //	btelpv30_17.wav
[18] tá hhh //	btelpv30_18.wav

The annotators who took part in the test correspond to the four annotators with the best results in the pilot test. As an annotator has to undergo extensive training before being able to participate in an annotation test like the one carried out in this study, we were unable to find participants other than those who also participated in the pilot test. However, we did make sure that the utterances used in the actual test had not been used in the preliminary one.

Before proceeding to showing and discussing the results, we must add that the participants were sent the protocol by email, along with the files that they should use in the annotation task. Finally, the task had to be carried out within one month, during which time a revision of the annotation should be conducted. The participants were not allowed to discuss any aspect of the annotation while they were doing the task.

5.2.2 Preparing the data to compute *k*

Each participant turned in the XLSX file he or she had been sent but with the target utterances fully annotated. In order to prepare the material for calculating the kappa coefficients, the tagged utterances were extracted and processed so that only the tags were kept (see the second column of Table 5.4). Furthermore, the illocutionary and non-illocutionary tags (COM, COB and NTP) were replaced with "XXX" (see Table 5.4), since we were interested in measuring the agreement only for the identification of TOP units. The reason for having asked the participants to distinguish the types of illocutionary units was related to possible future studies.

Table 5.4: Tags extracted to run the agreement test

Tagged utterance	**Tags**
[19] a essa altura /=TOP= já dá para ver claramente os sinais da febre do Glaciar Exploradores //=COM=	TOP XXX
[81] a tecnologia /=TOP= chegou pra ficar /=COB= ninguém tem dúvida //=COM=	TOP XXX XXX

Once extracted, the tags of all the four participants were organized in a data frame and then the kappa computations were done considering prosodic unit by prosodic unit. As in the

pilot test, the data was processed using *R* (R Core Team) and the computations were done using the *kappam.fleiss* function from the *irr Package* (GAMER; LEMON, 2012).

The results of the test are shown in the following section

5.3 Results

By providing the annotators with (a) the utterances embedded in their immediate environment, (b) the guidelines containing specifications as to what constitutes a TOP and how to detect it, and (c) banning the visualization of prosodic cues, the kappa coefficient calculated was $k = 0.79$, which represents a level of substantial agreement according to the benchmarks in Table 5.1 proposed in Landis and Koch (1977, p. 165). The table below shows the coefficients calculated considering the annotation of three participants at a time.

Table 5.5: *k* coefficients computed considering three annotators at a time

Annotators	**k**
B, C, D	0.81
A, C, D	0.79
A, B, D	0.78
A, B, C	0.78

The number of TOPs identified by each annotator ranged from 48 to 55 ($M = 52$, $SD = 3.2$). Full agreement was reached in 36 cases. In 11 cases, only three annotators agreed, and in 5 cases only two did. The number of TOPs that each annotator detected and that was not detected by any other annotator ranged from 2 to 8 ($M = 5.25$, $SD = 2.5$).

Considering the utterances from Natural Context, Media, and Telephonic separately, the results are as follows. For the Media section, whose 40 utterances added up to 234 prosodic units, the overall agreement achieved was substantial ($k = 0.80$), just as for the Natural Context section ($k = 0.79$), with its 240 prosodic units.

For the utterances from Telephonic section, on the other hand, the overall agreement was relatively lower ($k = 0.66$) but with k still falling within the range of values considered substantial, (i.e. 0.61–0.8, see Table 5.1). We have to consider that the number of utterances sampled from this section represents half of the number sampled from each of the other two sections. With a total of only 67 prosodic units, the 20 utterances from the Telephonic section had considerably fewer prosodic units (3.4, on average) than the other two sections (6.0, on

average). Moreover, the interactions recorded in the Telephonic section not favor the occurrence of TOP units. This is because TOP shows a higher incidence in interactions characterized by being less interactive, less dependent on the immediate context, and more complex in terms of textual structure. The telephone calls documented in C-ORAL-BRASIL II, however, are usually the opposite of that, as they are for the most part short and simple in terms of textual structure. Finally, the kappa statistic is affected by the prevalence of the phenomenon under consideration in such a way that a small kappa value does not necessarily reflect a low degree of agreement for rare findings (VIERA; GARRETT, 2005). It is important to take that into consideration when assessing the coefficient calculated for the utterances from the Telephonic section, in which no more than 2 TOP units were annotated by each participant.

Table 5.6 below summarizes the results discussed so far and shows the mean number of TOPs detected by the annotators in each section and in all of them considered together.

Table 5.6: Kappa coefficients (*k*), total number of prosodic units (IU), and mean number of TOPs detected in each section, and standard deviation (SD)

Section	**k**	**IU**	**Mean (SD)**
Media	0.80	234	28 (0.7)
Natural Context	0.79	240	22 (2.9)
Telephonic	0.66	67	1.8 (0.4)
General	0.79	541	52 (2.7)

5.4 Discussion

The results of the agreement test reported here are hardly comparable with similar studies conducted within other frameworks (see LÜDELING et al., 2016). One of the reasons for that, of course, is the incompatibility among definitions. First of all, it is often the case that the notions of information structure are defined in a way that they can be applied both to spoken and written language. This does not happen with L-AcT, which considers prosody and information structure as two inextricably related notions. In the specific case of TOP, we have seen that it is defined not only according to its function but also in prosodic and distributional terms. Moreover, the prosodic form of the unit is characterized by specific patters of f0 variation, duration, and intensity, not simply by a vague or loosely specified prosodic prominence. In addition, the aboutness relationship of TOP, according to L-AcT, is established with the illocution, while for other frameworks this relationship is eminently semantic, being

established between elements in a predication (see, e.g., CHAFE, 1976, LAMBRECHT, 1994, KRIFKA, 2008).

According to L-AcT, the identification of TOP does not depend on the novelty or givenness of its referent, nor does it depend on positing a syntactic relationship between the expression in TOP and that in illocutionary unit. This means that the annotator has to detect a well-defined prominence in a prosodic unit marked with a non-terminal break and then evaluate in from the semantic and pragmatic point of view.

The illocution is always new, in the sense that it is an unpredictable result of an unconscious drive or a pulsion, in Cresti's (2000) terms. The information status of referents designated by linguistic expressions in a prosodic unit is not relevant for determining whether the unit is TOP or not. But as the TOP unit supplies a domain of identification to highlight, specify or even reveal a certain referent according to which the illocution must be interpreted, its referent may have different degrees of accessibility and novelty.

To work with a definition of topic such as that proposed by Lambrecht (1994) is quite different from working with the definition adopted in this study. According to Lambrecht, an entity is the topic of a sentence if the proposition conveyed by the sentence can be pragmatically construed as being about that entity (LAMBRECHT, 1994, p. 120). In order to determine the topic of a sentence, one has often to be able to access the discourse context, given that Lambrecht's topic is an abstract notion that does not show stable formal correlates. And other approaches (not only Lambrecht's) are very different from L-AcT with respect to the way information structure is accounted for (see chapter 1), we can see that the results of the study reported in this chapter can hardly be compared to others carried out within different frameworks (see LÜDELING et al., 2016).

Resuming the discussion of aspects more directly associated with the outcomes of the present study, the analysis of the full disagreements in the annotation task suggests some possible sources of confusion. The first one is related to cases in which an annotated prosodic break was too subtle (or perhaps even absent), e.g. the bold part in the examples below.

(5.3) bmedsc04_1_94 (ex5.3.wav)
o mercado de trabalho / pra engenharia agrícola / é muito amplo //
the job market / for agricultural engineering / is very large //

(5.4) bmednw03_2_3 (ex5.4.wav)

o ministro da Fazenda / Guido Mantega / culpou o juiz americano / pelo calote argentino / que ele diz que não existiu //
the finance minister / Guido Mantega / blamed the American judge / for Argentina's default / which he claims did not happen //

On close inspection, the first prosodic break annotated in ex. (5.3) is barely noticeable. This utterance was present in the test, and because the annotators were not allowed to alter the prosodic annotation, they were forced to tag it. One of the annotators, disregarding the prosodic criterium and focusing on the semantic one (i.e. the ability of the unit to constitute a cognitive domain), ended up tagging it TOP. The same thing happened with the utterance in ex. (5.4), despite the fact that the prosodic break after the phrase "*o ministro da fazenda*" is even feebler than the one in the previous example.

There were also cases in which a unit that was eventually determined to be a SCA unit had been mislabeled TOP. This happened probably because the prosodic break in these cases delimits a coherent morphosyntactic structure that, from the semantic point of view, could indeed be a TOP unit. These cases are similar to examples (5.3) and (5.4), but they show clearly noticeable prosodic breaks. They correspond to 26.3% of the prosodic which were tagged TOP by only one annotator.

Additionally, one bound comment (COB) was mislabeled TOP. This probably occurred due to inattention on the part of the annotator, who seem to have misinterpreted the prosodic signal of continuation of the unit, which is characteristic of COBs in general, as a prosodic prominence associated with the function of TOP. The utterance in question is shown below in (5.5).

(5.5) bmedsp03_1_147 (ex5.5.wav)
então a gente tá vendo aí o Pato / porque é o centro das atenções essa semana aqui no Corinthians //
so we are looking at Pato over there / because this week he is the center of attention here at Corinthians [team's practice facility] //

We also counted one case in which the locutive content of the unit could would qualify as a dialogic unit or a TOP unit. As the expression *então* in ex. (5.1) discussed above, the expression *afinal* can either take up a referential meaning or fulfill a dialogic function. In ex.

(5.6), it appears to be non-referential, but it was labeled TOP by one annotator, who was likely motivated to do so because of the relative prosodic strength with which *afinal* is uttered, as can be appreciated by playing the corresponding sound file.

(5.6) bmedsc03_3_8 (ex4.7.wav)
afinal / quem não gosta / ou não gostaria de ter um carro //
after all / who doesn't like / or wouldn't like to have a car //

As can be seen, inattention on the part of the annotator regarding the semantic characteristics of the unit also appears to have caused error.

In a future study, it would be interesting to have a larger sample, particularly in the case of utterances from Telephonic section, which in this study were only 20. In addition, doing such a test with other information units may bring insights as to their appropriateness and reliability, thus allowing for refinements in some particular aspects of L-AcT.

6 The prosodic forms of the topic information unit

According to L-AcT, the topic information unit (TOP) is realized by specific prosodic patterns, characterized by defined melodic shapes that are accompanied by syllable lengthening and higher intensity on certain portions of the unit. Visual and auditory analyses of a large number of TOP units combined with *f0* manipulation experiments have suggested that this information unit has at least three prosodic forms. This seems to hold true for all languages in which the unit has been systematically studied thus far, namely Italian (FIRENZUOLI; SIGNORINI, 2003, SIGNORINI, 2005), Brazilian and European Portuguese (ROCHA, 2012, MITTMANN, 2012, respectively), and American English (CAVALCANTE, 2015, RASO et al., 2017).

A unit carrying out the information function of TOP is always composed of a *prosodic nucleus* (FIRENZUOLI; SIGNORINI, 2003; CRESTI, 2011), which encompasses no more than four or five syllables whose role is to impart the informational function of the unit. A TOP unit may also be composed of non-nuclear syllables. Their function, however, is restricted to the morphosyntactic and semantic level of the unit, and depending on whether they precede, intervene, or succeed the nuclear syllables, they are called *preparation*, *linking portion*, or *coda*.

Research on the information unit of TOP carried out within L-AcT, whose latest synthesis can be found in Raso et al. (2017), has identified at least three prosodic forms through which the unit can be realized. The description of these forms – referred to as Type 1, 2, and 3 – is centered on the prosodic nucleus, its melodic shape and the relatively higher duration and intensity values exhibited by the syllables that constitute it, as the following examples show.

(6.1)

Type 1: rise-fall *f0* contour aligned with the last stressed syllable of the unit (and post-stressed ones, if any)

afamdl01_ 80 (ex6.1.wav):
once I get my ex**perience** /=TOP= I'll be up there too / in the top-four salesmen //

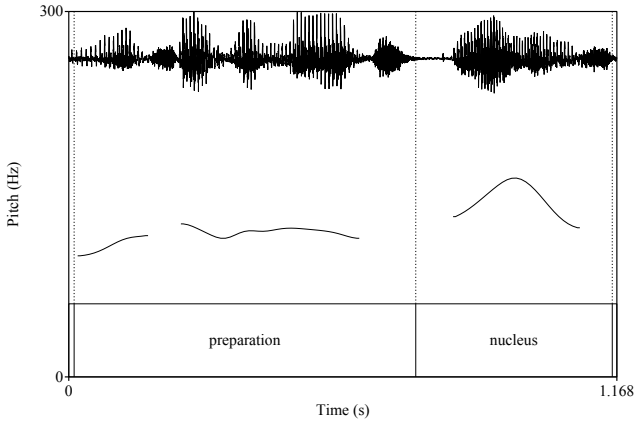


Figure 6.1 – *f0* contour and waveform of Type 1 prosodic form shown in ex. (6.1) above. The nucleus corresponds to the portion in bold in the example

(6.2)

Type 2: rising *f0* contour aligned with the last stressed syllable (and post-stressed ones, if any)

afamdl01_67 (ex6.2.wav):
but in a **sense** /=TOP= I need a [/1] some type of steady income //

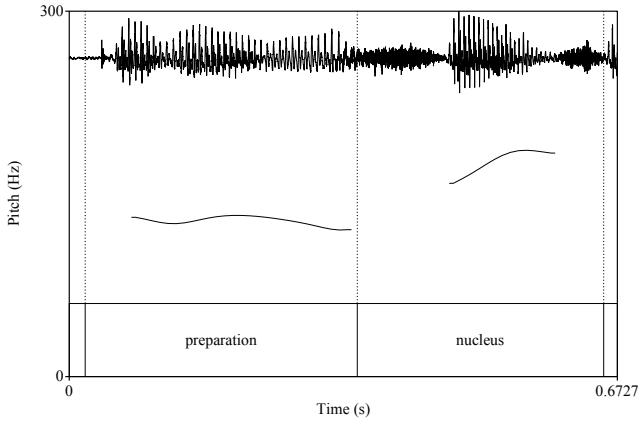


Figure 6.2 – *f0* contour and waveform of Type 2 prosodic form shown in ex. (6.2) above. The nucleus corresponds to the portion in bold in the example

(3)

Type 3: two often discontinuous semi-nuclei; the first one features high to extra-high *f0* values; the second features lower *f0* values combined with syllable lengthening; the non-nuclear syllables that frequently join the two semi-nuclei correspond to the aforementioned preparation.

afamcv04_138 (ex6.3.wav):

when **Mary** tells me to get **a sleep** *over the weekend* /=TOP= you know I need to get sleep over the weekend //

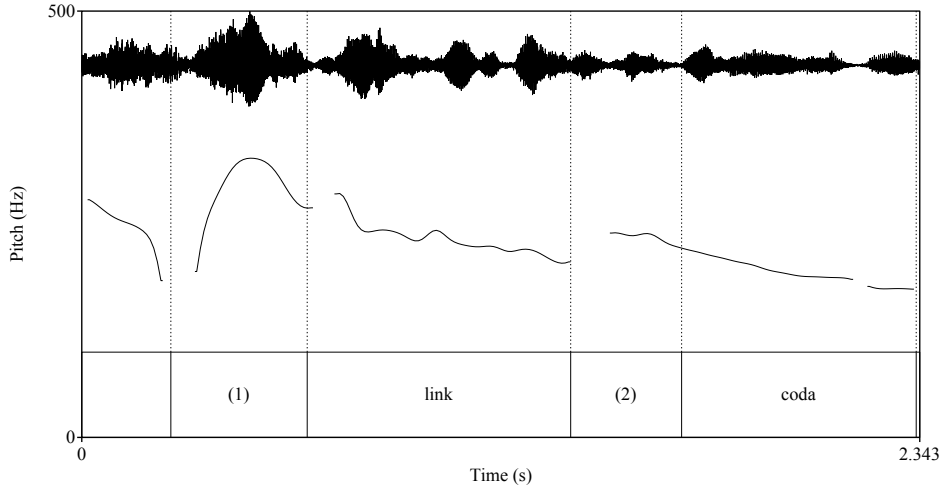


Figure 6.3 – *f0* contour and waveform of Type 3 prosodic form shown in ex. (6.3) above. The semi-nuclei are numbered and correspond to the portions in bold in the example, while the coda is signaled in italics

In Cavalcante (2015), it is suggested that TOPs found in American English appear to take yet another prosodic form, characterized by *f0* contours with no relevant movement and semi-nuclei that, much like Type 3, may be surrounded by non-nuclear syllables. This possibility will be further examined in this chapter. This prosodic form has been called *Flat form* and is shown in the bottom-right panel of Figure 6.4 below, which also shows diagrams for the other three forms.

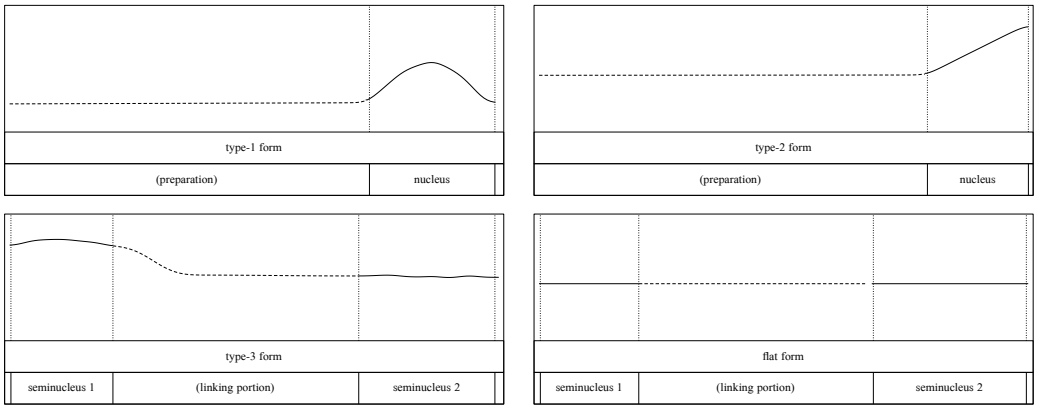


Figure 6.4 – Diagram of the prosodic forms of TOP

Although based on careful observation of data from spontaneous speech corpora, the descriptions presented above require further investigation, as they have not been validated statistically. Therefore, the aim of this chapter is to refine and validate the descriptions of the prosodic forms of TOP. This will be carried out with data collected from corpora of Italian (IT), European Portuguese (EP), Brazilian Portuguese (BP) and American English (AE). We will use *Functional Data Analysis* along and *Functional Principal Component Analysis* to evaluate the classification of the prosodic forms proposed in previous studies (see RASO et al., 2017) and to derive statistical models for the different realizations of the unit. In addition to that, we aim to look more closely at the case of Flat TOP units, which were first reported in Cavalcante (2015) as units that exhibit very little *f0* movement, but which nonetheless cause a clear impression of pitch variation.

6.1 Functional Data Analysis

The core statistical approach utilized in this study is *Functional Data Analysis* (FDA), a term coined by Ramsay and Dalzell (1991) which refers to methods to display and analyze time-varying signals (RAMSAY; SILVERMAN, 2005). It is a relatively new set of techniques that has been applied to different sorts of data and, within linguistics, to the study of different dynamic phonetic phenomena.

The overall goals of FDA, according to Ramsay and Silverman (2005, p. 9), are:

1. to represent data so as to assist further analysis (both quantitative and qualitative)
2. to display data highlighting common patterns and features
3. to study sources of patterns and variability in the data
4. to account for variation in a dependent variable on the basis of independent variable information.

One of the many interesting aspects of FDA is that it allows one to unveil patterns independently of prior expectations that a researcher may have. For linguistic studies, Lucero and Koenig (2000) suggest that FDA may be an optimal tool to normalize voice signals, given that the technique keeps the relevant characteristics of the original signals without compromising the analysis of local irregularities individually.

Lee et al. (2005) utilize FDA to examine the articulatory timing and its interplay with prosodic aspects, showing how useful FDA may be to detect time patterns in linguistic experimentation. Koenig et al. (2008) utilize FDA to study the production of fricatives by children and adults, decomposing age-related variability into indices of amplitude and temporal (phasing) variations. Gubian et al. (2009) illustrate how FDA, coupled with *Functional Principal Components Analysis* (F-PCA), may be utilized for the study of transitions in speech signals. The researchers applied FDA in a study of the reduction of French *c'était* (i.e. /se'tɛ/ to [stɛ]) in conversation, seeking to establish whether the process is gradual or categorical. The results of their study, which is part of a broader project dedicated to understanding the role of fine phonetic details in speech comprehension and production, suggest that FDA is indeed a particularly powerful tool for the analysis of dynamic phonetic detail as well as the dynamics of other aspects related to speech.

In studies such as the present one, FDA can be employed as a means to contribute to a better understanding of data characteristics that are already known as well as to possibly unveil new features and patterns. We will examine *f0* data extracted from corpora of IT, EP, BP and AE, which are described in chapter 4. In particular, we will look at *f0* contours associated with the realization of the information function of TOP so as to verify the appropriateness of the descriptions found in the specialized literature, seeking to refine these descriptions and possibly reveal features in the contours that the visually-mediated analyses conducted thus far have not allowed to detect.

6.1.1 FDA procedures

Fundamentally, applying FDA involves two preliminary steps: *data smoothing* and *data registration* (also referred to as *time-alignment*). In the data smoothing phase, discrete data points are converted into a linear combination of smoothly varying functions, e.g. the B-splines used by Gubian et al. (2009) as well as in the present study. The ultimate goal of data smoothing is to highlight patterns in the signals by reducing short-term deviations that may arise due to measurement errors, signal noise, and micro-melodic phenomena, which are of no interest to our investigation. The line fit through the points in Figure 6.5 below illustrates the data smoothing of a Type 3 unit from the AE minicorpus using 10 B-splines.

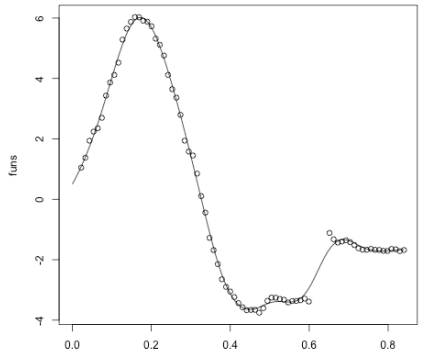


Figure 6.5 – Smoothed *f0* contour of a Type 3 TOP

The second step, called *data registration*, involves eliminating the temporal mismatch among input signals by normalizing them on a same time interval (hence the alternative term, *time-alignment*). As Lee et al. (2005) put it, the aim of time alignment is to find a common time path among signals with different properties by expanding and compressing (i.e. warping) the time of a signal against a reference by means of a time warping function.

In this phase, it is possible to take into account prior knowledge about the phenomenon under investigation by means of a process called *landmark registration*, whereby time-registered versions of the functions are produced in such a way that their *landmarks*, i.e. points in the signal deemed relevant for the phenomenon being studied, coincide in time. Figure 6.6 below shows the landmark registration in some Type 3 curves from the data analyzed in this study. The dots, coded in different colors, mark the semi-nuclei of the curves. The first semi-nucleus is delimited by the red and green dots, and the second, by the dots in different shades of blue.

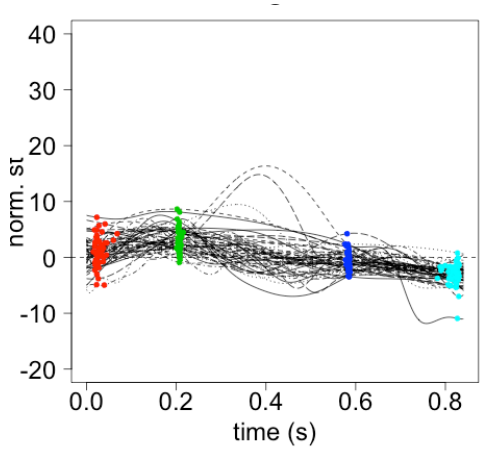


Figure 6.6 – Landmark registration of Type 3 *f0* curves (see below for discussion)

Once the smoothing and registration steps outlined above have been done, the analysis proper may be conducted, for which different multivariate statistics techniques may be used (GUBIAN et al., 2009; RAMSAY; SILVERMAN, 2005). F-PCA was the technique utilized in the present study. According to Ramsay and Silverman (2005, p. 147), F-PCA constitutes a key method for the analysis of functional data, as it provides means for obtaining and displaying the variation of a set of multidimensional data.

6.1.2 Functional Principal Component Analysis

Traditional Principal Component Analysis (PCA), which is meant to be applied to scalar data, constitutes a set of statistical techniques for unveiling the main modes of variation in a set of multi-dimensional data and also for displaying those modes. F-PCA, the functional counterpart of the technique, is used with similar goals, but it meant to be applied to functions rather than scalars.

According to Gubian et al. (2014), F-PCA can be employed to obtain a linear model *f(t)*[45] of a set of curves after they have been subjected to the FDA smoothing and registration processes. The model *f(t)* is obtained by adding together the mean curve $\mu(t)$ and a small number of Principal Components curves $PC_i(t)$ that are computed on the basis of all the input curves and that represent different deformations of $\mu(t)$. Furthermore, for each input curve, F-PCA specifies a set of weights s_1, s_2, s_3 etc., referred to as *PC scores*, which, multiplied by the corresponding curve $PC_i(t)$, produce the best approximation *f(t)* of a given input curve, according to the equation in (6.1). The number of PCs vary depending on the data under analysis. In our case, as will be shown below, the first 2 components account for over 80% of the variance and thus suffice for modeling each type of prosodic form.

$$f(t) \approx \mu(t) + s1 \cdot PC1(t) + s2 \cdot PC2(t) + s3 \cdot PC3(t) \ldots \quad (6.1)$$

Gubian et al. (2014) uses the following cooking analogy to clarify the job done by F-PCA. Each smoothed and time-aligned curve that serves as an F-PCA input corresponds to a different dish. The aim of the technique is thus to establish a small number of ingredients, corresponding to the principal components curves $PC_i(t)$, that can be used to best reproduce the

[45] Here the input curves are *f0* traces obtained with Praat and subjected to FDA.

individual dishes (again, the input curves) by varying the amount of each ingredient alone. The amount of each ingredient corresponds to the weights, or PC scores, referred to above.

Figure 6.7 below shows the input curve of a Type 1 nucleus (dashed line) and its approximation (in blue) calculated according to equation (6.1) reproduced below as (6.2) with the coefficients s_1, s_2, and s_3 plugged in.

$$f(t) \approx \mu(t) + 0.8067 \cdot PC1(t) + 0.0397 \cdot PC2(t) - 0.4905 \cdot PC3(t) \quad (6.2)$$

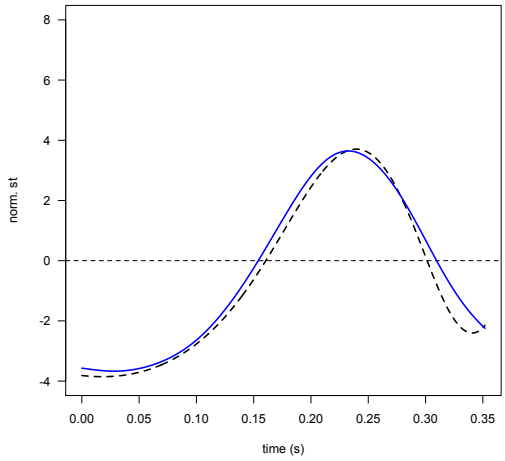


Figure 6.7 – A Type 1 nucleus curve approximation

In sum, F-PCA provides a model of a set of functional data consisting of the mean curve $\mu(t)$ and the principal components curves $PC_1(t)$, $PC_2(t)$, $PC_3(t)$ etc. computed for all the curves in the data, together with the PC scores s_1, s_2, s_3 etc. computed for each input curve individually. The PC scores can be used in further statistical analysis as well. In the present study, we used ANOVA to assess the classification scheme proposed in previous studies (see Raso et al., 2017) by testing the separability of a number of curves associated with the different types of prosodic forms of TOP proposed in previous studies.

The PC scores are calculated without any class membership information, but they may be used for establishing quantitative associations among categories, e.g. the prosodic forms analyzed in the present study. Thus, class membership of individual curves and their scores can be plotted so as to find out whether they are coherently associated. The scatterplot in the Figure 6.8 below, showing the PC_1 and PC_2 values of curves obtained from the AE minicorpus and classified according to the types presented above, illustrates this sort of class analysis.

As can be seen, the plot suggests that the classification proposed in previous studies (see Raso et al., 2017) and presented in the beginning of this chapter may be correct, despite the presence of some class overlapping and a few outliers, which will be dealt with below.

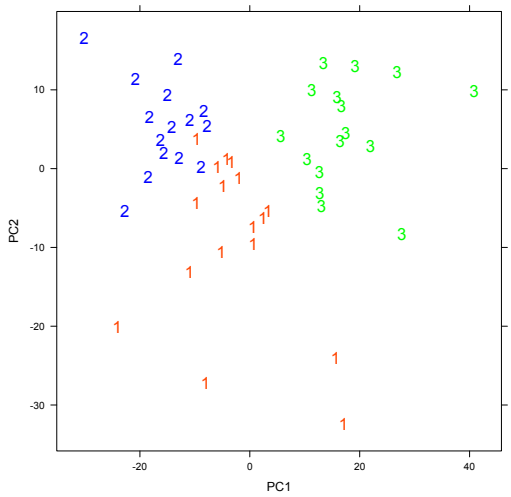


Figure 6.8 – Scatterplot of PC_1 and PC_2 corresponding Types 1, 2, and 3 curves of TOP. The digits 1, 2 and 3 specifies the prosodic type of the curve corresponding each data point in the plot

The validity of the classification will be quantitatively assessed using data from different languages and applying ANOVA to the numerical descriptors s_i obtained with F-PCA. The qualitative analysis will be carried out by producing models using the equation in (6.1) and comparing them to the descriptions of each type of prosodic form found in previous studies (see Raso et al., 2017).

In the following section, the data used in our analysis will be presented.

6.2 The data

The data used in this study come from corpora compiled according to the methodology established by L-AcT for the study of spontaneous speech. The corpora used encompass the AE and BP minicorpora, the EP component of the C-ORAL-ROM and a portion of what will make up the new version of the IT minicorpus (for details about these resources, see chapter 4).

Cavalcante (2015) conducted an analysis of TOP units from the AE minicorpus. However, we have carried out a thorough revision of the minicorpus in preparation for the present study. Following the revision, which involved both the prosodic segmentation and the annotation of information functions, all the TOP units were reexamined through visual and

auditory inspection and reclassified according to type of prosodic form. Out of the 446 TOPs found in the minicorpus, 296 were deemed suitable for a reliable classification. Unlike Cavalcante (2015), where all TOPs were classified, we chose to err on the side of caution and focus only on those units which exhibit a certain level of acoustic quality. The reasons for not using some of the TOP units found in the minicorpus include:

- presence of overlapping speech, a phenomenon that is particularly common in dialogues and conversations
- presence of background noise interfering with *f0* readings
- low level of acoustic quality of some texts
- occurrence of a given TOP unit in an interrupted utterance when the absence of the illocution hampered an adequate classification
- short syllabic extension, which oftentimes makes it extremely difficult even to determine whether the unit is actually a TOP or not
- presence of scanning phenomena separating the two semi-nuclei of Type 3 TOPs.

In addition, Lists of Topics (TPLs)[46], with only 10 occurrences in the minicorpus, were not subjected to FDA. Table 6.1 shows the distribution, according to type, of the TOP units selected from the AE minicorpus.

Table 6.1: Distribution per type of TOPs from the AE minicorpus

Type	**Frequency**	**%**
1	23	7,8
2	23	7,8
3	224	75,7
Flat	26	8,8
Total	296	100,0

Before proceeding, it is important to stress that the proportion of each prosodic type presented here is different from the frequency reported in Cavalcante (2015), who classified 398 TOPs following less restrictive criteria for the selection of curves to be classified. This is due mostly to our decision to be much more cautious in our selection of units. Since our criteria

[46] See chapter 2 for details about TPLs.

were established without aiming to keep the same proportions of types fount in Cavalcante's (2015) sample, the difference was expected. In addition to that, we classified a few TOP units different from how Cavalcante (2015) did.

Not all of the 296 classified TOP units were used in the FDA. The sample used in the analysis was decided based on the number of the least frequent forms, i.e. Types 1, 2 and Flat. The latter type, at this point, cannot be considered a class in its own right, and this will be discussed below. Table 6.2 shows the number of TOPs used in the crosslinguistic comparison.

Table 6.2: TOP units used in the crosslinguistic analysis

Language	**Type 1**	**Type 2**	**Type 3**	**Flat**	**Total**
AE	14	15	14	24	67
EP	1	12	17	1	31
BP	11	20	20	4	55
IT	8	10	10	2	30
Total	34	57	61	31	183

Besides the TOP units computed in the table above, another 43 Type 3 units from the AE minicorpus were examined in the part of the analysis that uses landmarks.

Regarding the TOP units of the other languages, the situation is as follows. From the BP minicorpus, 55 out of the 100 TOP units analyzed in Mittmann (2012) were selected based on acoustic quality and then subjected to a revision. Four of the selected units turned out to be reclassified as Flat. More will be said about that below.

As discussed in chapter 4, the EP data was taken from a sample created by Rocha (2012), since there is no informationally tagged corpus of this variety of Portuguese. The data consists of 72 TOP units carefully selected by means of visual and auditory inspection of the EP component of C-ORAL-ROM (CRESTI; MONEGLIA, 2005). In the present study, we analyze 31 of those TOP units, which were selected based on the acoustic quality and then checked for possible misclassification.

Finally, the IT data comes from the new version of the IT minicorpus, which is currently being prepared by members of the LEEL team (BOSAGLIA; RASO, in preparation). Like its first version, this minicorpus will be comparable with the BP and AE minicorpora. At the time the analyses reported here were carried out, only the dialogues and conversations were available. From them, we selected 30 TOP units based on acoustic quality, and then classified them.

6.2.1 Preparing the audio files

For the preliminary phase of the analysis, we used Praat (BOERSMA; WEENINK, 2019) to annotate the boundaries around the nuclei of the units, which were used as landmarks in the pilot FDA. Two different Praat scripts were written to obtain and store the time-stamped *f0* contours and the landmarks (see Appendices D and E). The FDA and F-PCA computations were carried following Gubian et al. (2015) using a script written in R (R Core Team, 2019) that can be obtained online[47] and adapted to be used with other kinds of data.

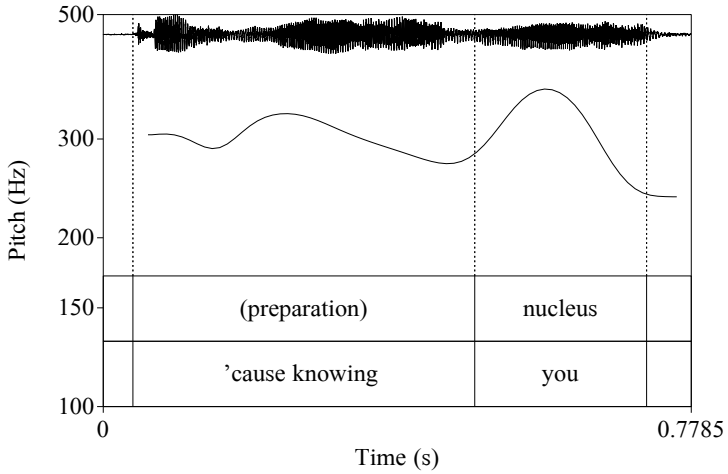


Figure 6.9: Type 1 *f0* curve with preparation and nucleus annotated

6.3 Preliminary analysis

Type 1 and Type 2 prosodic forms share two important characteristics: their nuclei occur at the end of the unit without any intervening syllables, unlike Type 3, which can (and often is) realized by discontinuous semi-nuclei. Furthermore, Types 1 and 2 nuclei admit a certain number of syllables preceding the nucleus constituting the aforementioned preparation, whose role is restricted to carrying the syntactic-semantic content of the unit (e.g. "cause knowing" in Figure 6.9). From an informational standpoint, the syllables that make up preparations – as well as linking portions and codas, for that matter – are "inert". Due to the similarity between these

[47] The FDA script used here other FDA resources are made available by the Gubian and colleagues at: http://lands.let.ru.nl/FDA/FDA_Papers.htm (accessed December 2019).

two forms, we decided to begin the analysis only with them. Due to statistical constraints, which require a similar number of cases in each class, we used 23 units of each type.

6.3.1 Sorting out Type 1 and Type 2

We began by subjecting the entire tone unit of each TOP to FDA. However, since the non-nuclear portions of the units are quite heterogeneous, our initial analysis suggested that the two forms were not separable. The registration of these curves is shown in the figure below.

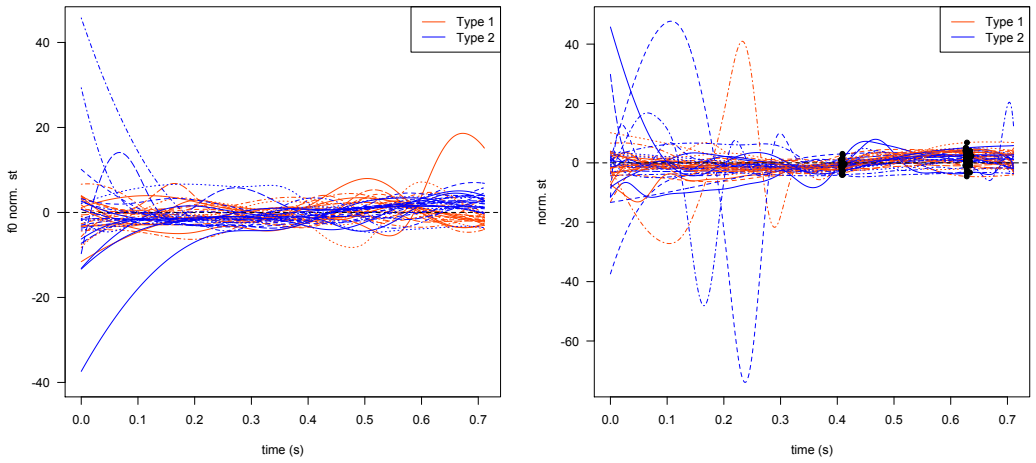


Figure 6.10 – Linear registration (left panel) and landmark registration (right panel) of Type 1 and Type 2

As seen in Figure 6.10, neither the linear nor the landmark registration of the curves suggest a common pattern. Assuming that the heterogeneity of the preparation portions of the curves were hindering the nuclear patterns from emerging, we decided to do the analysis once again, but at this time considering only the nuclear portions of each unit.

This time common patterns did emerge, as shown in Figure 6.11. The results suggest that the curves are actually separable according to their types, and it seems reasonable to consider them as forming two different classes, thus giving credence to the results from previous studies (see RASO et al., 2017). Figure 6.11 shows the registered nuclei curves corresponding to Type 1 and Type 2.

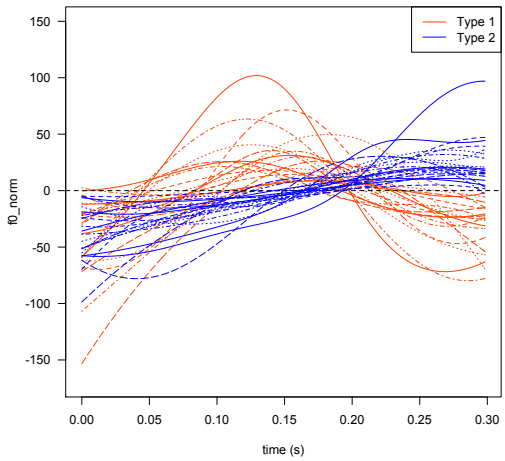


Figure 6.11 – Registered Type 1 and 2 nuclei

The registration shows that curves associated with a given type exhibit a common configuration. Type 1 nuclei (in red) tend to have a peak around their midpoint (somewhere between 0.1 and 0.2 s), while Type 2 nuclei exhibit an overall rising tendency starting at the beginning and going all the way till the end of the curves.

By performing F-PCA on the FDA output, the separability of the two forms is further supported. The mean curve $\mu(t)$ along with the curves $PC_{1\text{-}3}(t)$, which express the main modes of variation in the data, are shown in the four panels of Figure 6.12.

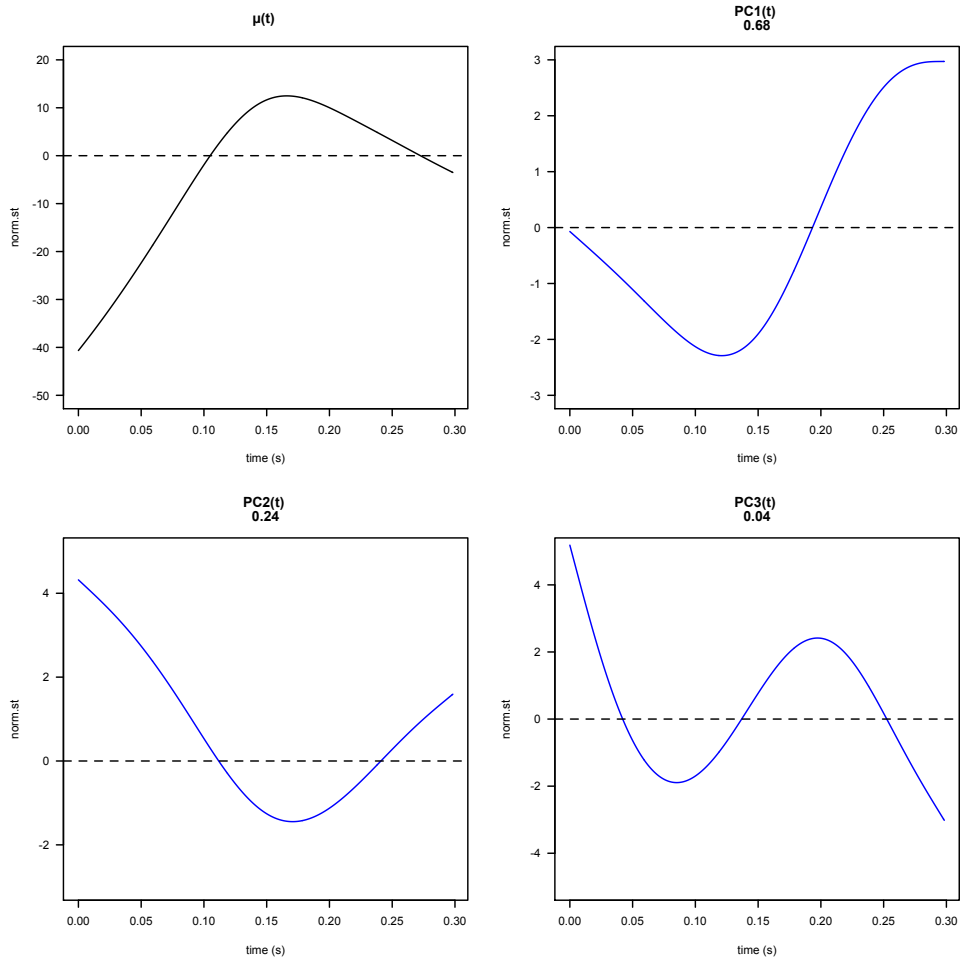


Figure 6.12 – $\mu(t)$, $PC_1(t)$, $PC_2(t)$, and $PC_3(t)$ computed for Type 1 and 2 nuclei, along with percentage of variability accounted for by each $PC_{1\text{-}3}(t)$ (0.68, 0.24, and 0.04 respectively)

The first three curves $PC_i(t)$ together account for 96.0% of the variability in the data. These curves are the "ingredients" that can be used to approximate each input curve. More interestingly, though, is the fact that they can be used to generate models of curves taken as a group, as will soon be shown.

A Kruskal-Wallis rank-sum test performed on the PC scores $s_{1\text{-}3}$ shows a statistically significant difference between the s_1 scores of each type (chi-squared = 29.4, df = 1, p-value = 5.8×10^{-8}). As for s_2 and s_3, there is no significant difference between them (s_2: chi-squared = 0.1, df = 1, p-value = 0.73; s_3: chi-squared = 1.08, df = 1, p-value = 0.3).

Figure 6.13 below shows the PC scores s_1 and s_2 for all the curves that make up the data under consideration.

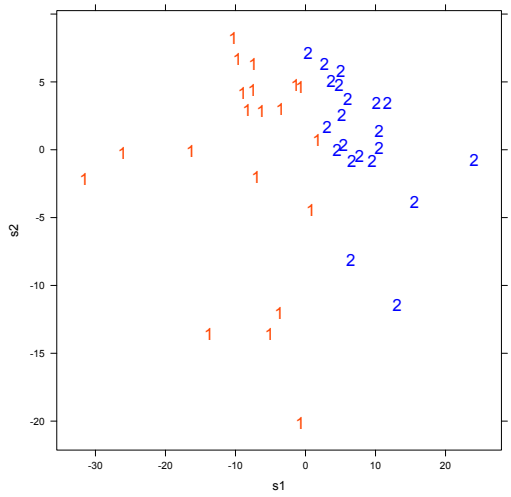


Figure 6.13 – PC scores s_1 and s_2 of Type 1 and Type 2 nuclei

Regarding the Type 1 curves exhibiting extreme s_2 values in the figure above, they are simply nuclei which have particularly pronounced rise-fall movements. Figure 6.14 shows the smoothed and time-aligned version of a Type 1 curve with s_2 = -12.13. A strongly negative s_2 value such as this turns $PC_2(t)$ upside down and "reinforces" its shape, which is coherent with the rise-fall movement characteristic of Type 1.

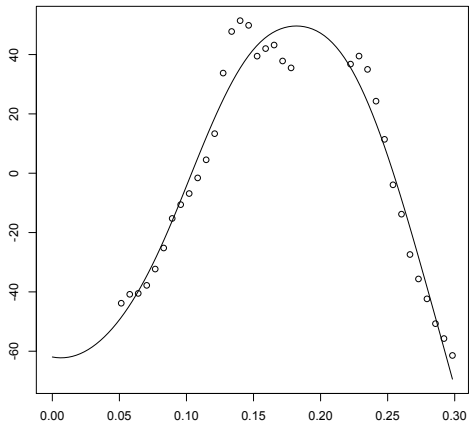


Figure 6.14 – Type 1 form with an extreme weight for $PC_2(t)$ (s_2 = -12.13)

In order to model the classes represented in this sample of Type 1 and Type 2 forms, we can use the median of s_1 scores – which, as shown above, give identity to each of these two classes – as a weight to adjust $PC_1(t)$, and then add together the curve $\mu(t)$ and the adjusted $PC_1(t)$ (see the left upper panel in Figure 6.12 as well as eq. (1)).

Since Type 1 curves have mostly negative s_1 values, once multiplied by their median, $PC_1(t)$ will be inverted. As for Type 2 curves, which have mostly positive s_1 values, the curve $PC_1(t)$, once multiplied by the median s_1 of Type 2, will have its rising movement reinforced.

By plugging in eq. (1) the median s_1 value computed for Type 1 and Type 2 separately (-7.3 and 6.5 respectively), we obtain the curves shown in Figure 6.15.

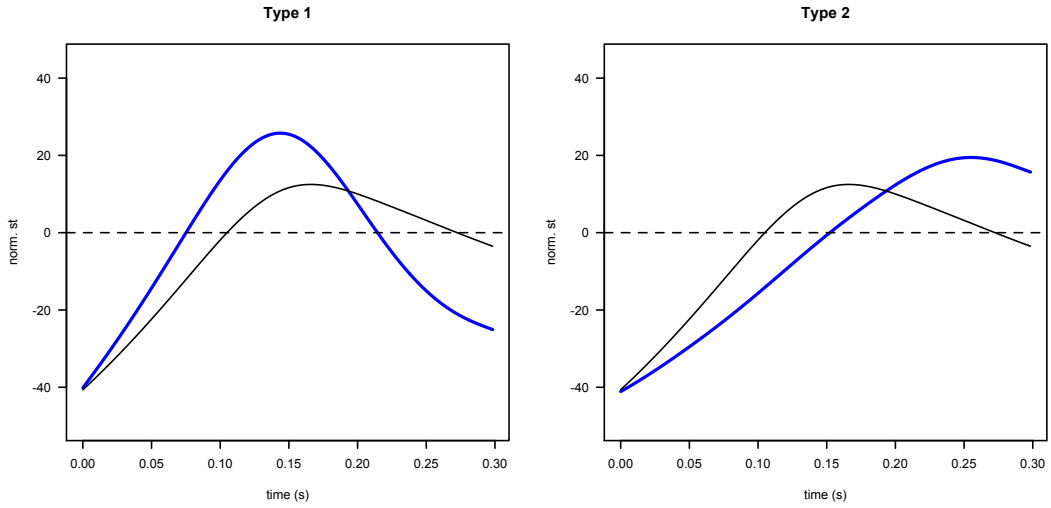


Figure 6.15 – Models produced according to $f(t) = \mu(t) + s_1 \bullet PC_1(t)$, where $\mu(t)$ is the mean curve (in black) and $PC_1(t)$ is the curve of the first component (see Fig. 5.12). The s_1 weight corresponds to the median s_1 value of the curves of each class. Left panel: Type 1(s_1 = -7.3); right panel: Type 2 (s_1 = 6.5)

The two shapes obtained through the procedure that has just been described closely resemble the melodic shapes of Type 1 and Type 2 nuclei outlined in the beginning of this chapter. We consider that this is an indication that previous studies (see RASO et al., 2017) were describing these two forms in an appropriate manner. But we still have more analyses to carry out before presenting a conclusion. After all, so far, we have examined only two of the four types of prosodic form that have been shown to exist, and in one language alone. In the next section we begin to tackle the curves of Type 3.

6.3.2 A first approach to the Type 3 prosodic form

The Type 3 prosodic form is unique in terms of how its nucleus can be realized. As mentioned earlier, this form is characterized as having two semi-nuclei that are often linked to each other by a few non-nuclear syllables, the so-called linking portion referred to above. The first semi-nucleus features high to extra high *f0* values, while the second one feature much lower *f0* values. Despite the formal difference, Type 3 fulfills the same informational function that is fulfilled by Type 1 and Type 2, and no study has thus far noted that a change in prosodic form implies a change in informational function.

The aim of the part of the analysis reported now was to determine common *f0* patterns in Type 3 curves. Apart from the different *f0* levels of each semi-nucleus of the form, it has been pointed out that they may show different movements, particularly the second semi-nucleus. In addition to that, Type 3 may have different number of syllables, with a few Type 3 forms having one or two syllables only, but others having many more. This heterogeneity constitutes a challenge for the analysis, and it may even cast some doubt as to the consistency of class.

For the statistical analysis of Type 3 curves, we considered the portion of the curves corresponding to segment delimited by the vowel onset of first semi-nucleus and the end of the last vowel of the second semi-nucleus. Thus, preparations (non-nuclear syllables preceding the first semi-nucleus) and codas (non-nuclear syllables following the second semi-nucleus) were not subject to FDA.

We began the analysis by looking at the landmark registration of 22 Type 3 curves from the AE minicorpora. In Figure 6.16, we can see something of the expected pattern: first semi-nucleus at a higher *f0* level (see the portions between the red and green dots) and second semi-nucleus at a lower level (the portions between the dots in different shades of blue). Furthermore, the first semi-nuclei of the curves seen in the figure tend to exhibit a rising movement, while the second tend to exhibit a falling one.

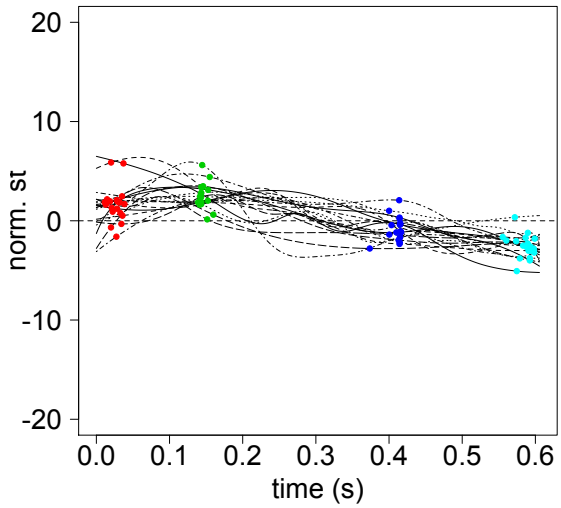


Figure 6.16 – Landmark registration of 22 Type 3 curves

We decided to apply the technique to a larger number of curves, taking advantage of the fact that there are many Type 3 units in the AE minicorpus. Figure 6.17 shows the landmark registration of 46 Type 3 curves, basically revealing a pattern akin to the one described by the curves in Figure 6.16, i.e. the first semi-nuclei with *f0* values above the mean and the second

semi-nuclei with lower values. Apart from that, there is a lot of variability in the data, both when it comes to the linking portions and the portions between the landmarks. As previously suggested, this variability was expected and could not be controlled for, given that we were dealing with naturally occurring data.

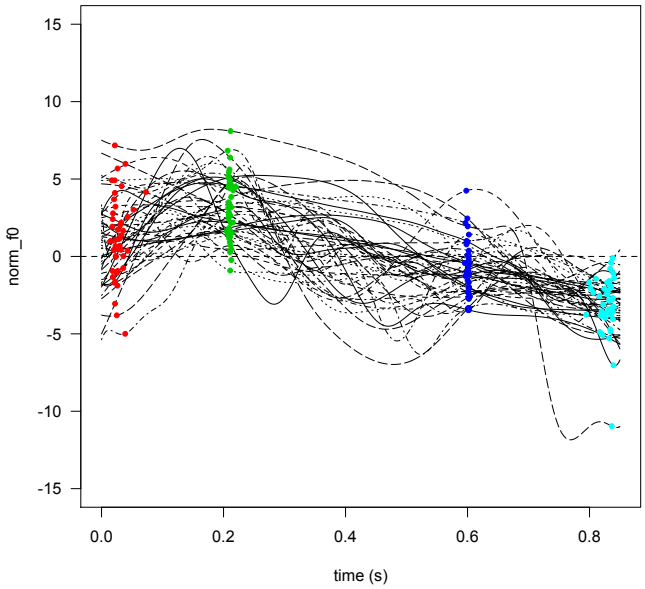


Figure 6.17 – Landmark registration of 46 Type 3 curves having at least 5 syllables each

The data in Figure 6.17 contain only Type 3 forms with five or more syllables. In part, the variability that they exhibit are accounted for by the sheer number of syllables that they have, since not only do they have two semi-nuclei – each of which exhibiting on average two syllables –, but some of them also feature portions that are optional from the informational standpoint. Compare them, for example, with the 23 Type 3 curves in Figure 6.18, of which no one contains more than four syllables.

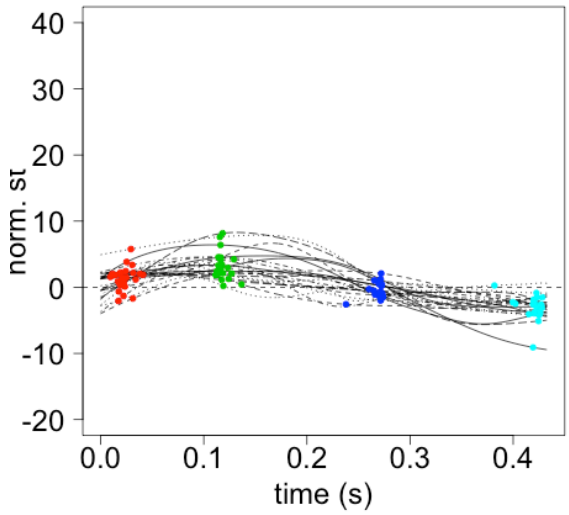


Figure 6.18 – Landmark registration of 23 Type 3 curves with no more than 4 syllables

As seen in Figure 6.18, both long and short curves generally exhibit the same pattern: a rising first semi-nucleus with high *f0* values and a falling second semi-nucleus ending at a level mostly below the mean. The smoother linking portion between the semi-nuclei of short Type 3 curves, which contrasts with the large variability of the linking portions of the longer Type 3 curves, can be attributed to their short size, which leaves little room for the occurrence of other phenomena.

The preliminary examination of Type 3 discussed in this section suggests that the classification carried out through visual and auditory means may indeed be consistent. Also, the landmark registered curves point to common tendencies that seem to hold for most Type 3 curves, both short and long. More will be said about Type 3 below, as we will compare it to Type 1 and Type 2 and establish their separability. But before doing so, we will look at the Flat prosodic form.

6.3.3 Flat prosodic units

First found in the AE minicorpus, some prosodic units functioning as TOP exhibit *f0* curves that are mostly flat, but which nonetheless yield the perception of pitch variation. These TOP units resemble Type 3 in that they, too, can be realized by discontinuous semi-nuclei. They have been referred to as Flat TOPs, as previously mentioned, and were considered a special kind of the Type 3 by Cavalcante (2015). The example and figure below show one such unit.

(6.4) afamdl05_57 (ex6.4.wav and ex6.4a.wav):

so / **the first hand /=TOP=** everyone has a club / so that they can't / discard a heart //

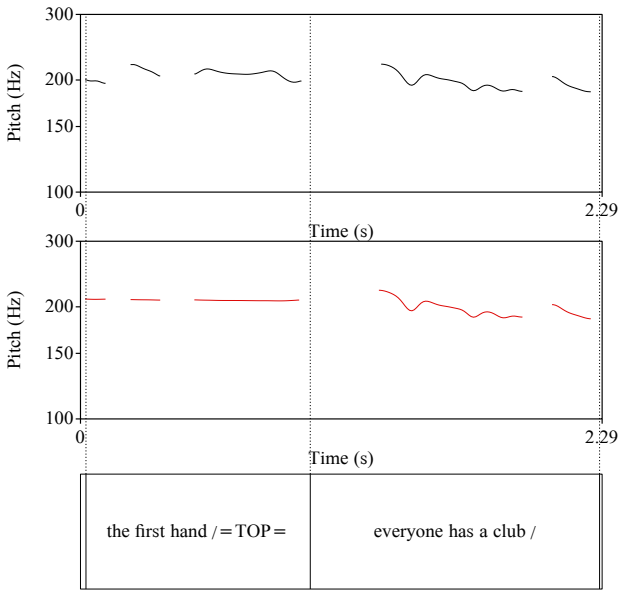


Figure 6.19 – *f0* contour of the Flat TOP in ex. (6.4). Original curve in the upper panel and manipulated curve in the middle

As can be seen in Figure 6.19, the unit in question is mostly flat, although we can hear some pitch variation in it. Interestingly, if the *f0* curve is flattened completely, as shown in the second panel of Figure 6.19 (see ex6.4a.wav[48] in the accompanying folder), the unit retains its ability to function as TOP, and the pitch variation effect is maintained as well. We still do not have a convincing explanation for this phenomenon, but we consider that an interplay of different acoustic parameters, like syllable lengthening and intensity, underlies in.

Flat TOPs are not very common, and we have found in our data 24 instances in AE, 4 in BP, 2 in IT, and 1 EP. The globally flat *f0* pattern of such TOPs can be seen in Figure 6.20, which shows the landmark registration of 14 Flat TOPs from the AE minicorpus, where they have first been found.

[48] In *ex6.4a.wav*, only the TOP unit has been manipulated.

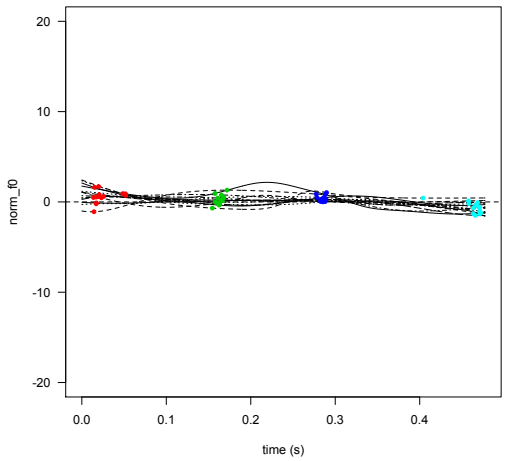


Figure 6.20 – Landmark registration of 14 flat TOPs from the AE minicorpus

As can be seen, the *f0* values drawing these curves are very close to the mean. Flat first semi-nuclei, corresponding to the portions between the red and green dots, may have a falling or a rising movement. The second semi-nuclei, however, is usually falling. These movements do not seem to be necessary, which is supported by the fact that these curves retain their functionality even when they are completely flattened and by experiments involving local manipulation, as can be appreciated by comparing *ex6.4.wav* and *ex6.4a.wav*.

The following section deals with the analysis of the four classes of prosodic forms discussed thus far, i.e. Type 1, Type 2 , Type 3, and Flat.

6.3.4 Class analysis: examining the four classes

As mentioned above, the outcomes of the statistical analyses with FDA and F-PCA can be used to validate the classification of the curves conducted manually. This has already been done in section 5.4.1, where we have shown that the PC_1 scores calculated for the Type 1 curves are statistically different from those calculated for Type 2 curve.

In this section we will do the same thing, but, this time, considering the three well-established forms – i.e. Type 1, Type 2, and Type 3 – as well as Flat, which was discussed in the previous section. Moreover, we will extend the analysis not only to AE data but also to data in IT, EP, and BP. We will start looking at the linear registration of 183 curves representing the four prosodic forms shown in the figure below.

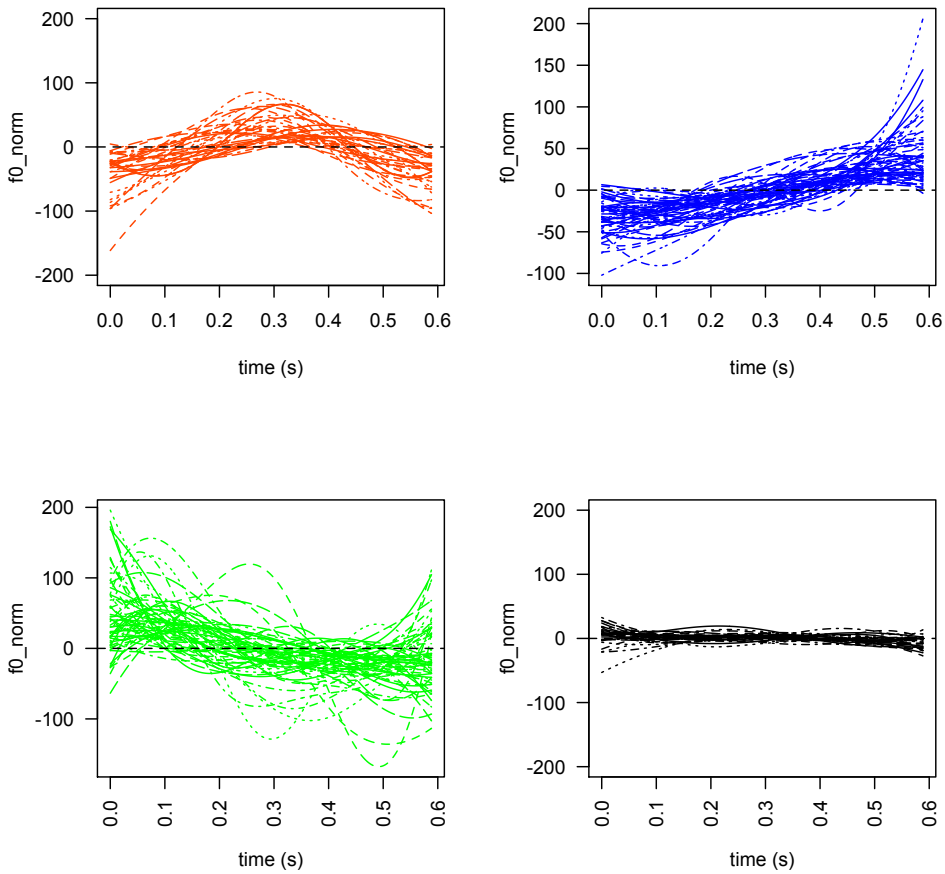


Figure 6.21 – Curve registration: Type 1 (upper left panel), Type 2 (upper right panel), Type 3 (lower left panel), and Flat (lower right panel)

By examining Figure 6.21, we can see that curves of the same type tend to cluster and exhibit a similar pattern. Type 1 curves (upper left panel) show the already the expected peak around their midpoints, while Type 2 curves (upper right panel) show the typical rising movement. These characteristics have been discussed in section 6.3.1, where Types 1 and 2 were examined, as well as in the beginning of the chapter.

As for Type 3 curves (lower left panel), they, too, show the expected pattern discussed in section 6.3.2 and outlined in the beginning of the chapter, namely a higher initial portion mostly above the mean and a lower end below the mean. Flat curves (lower right panel), though not perfectly flat, do not deviate much from the mean. In addition, they exhibit the pattern typical of Type 3 curves, with their initial portions – corresponding to the first semi-nuclei – usually at level higher than the final portion. The scatterplot in Figure 6.22 shows the scores s_1 and s_2 relative to the curves in Figure 6.21 above.

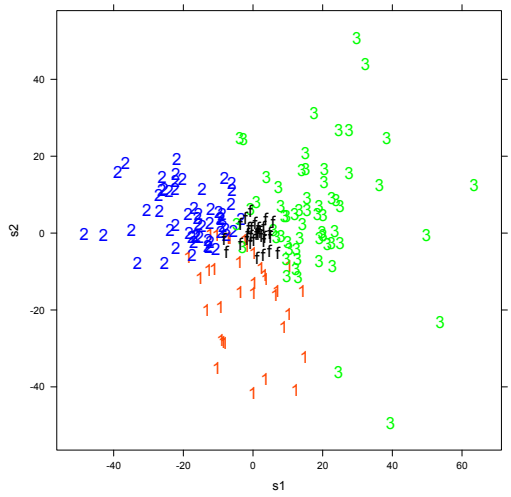


Figure 6.22 – s_1 and s_2 relative to the curves in Figure 6.21

The figure above shows that curves of the same type exhibit similar s_1 and s_2 values, suggesting that s_1 scores distinguish Type 1, Type 2, and Type 3 from one another, and s_2 scores distinguish Type 1 from the other classes. The same information can be seen in the two boxplots shown in Figure 23.

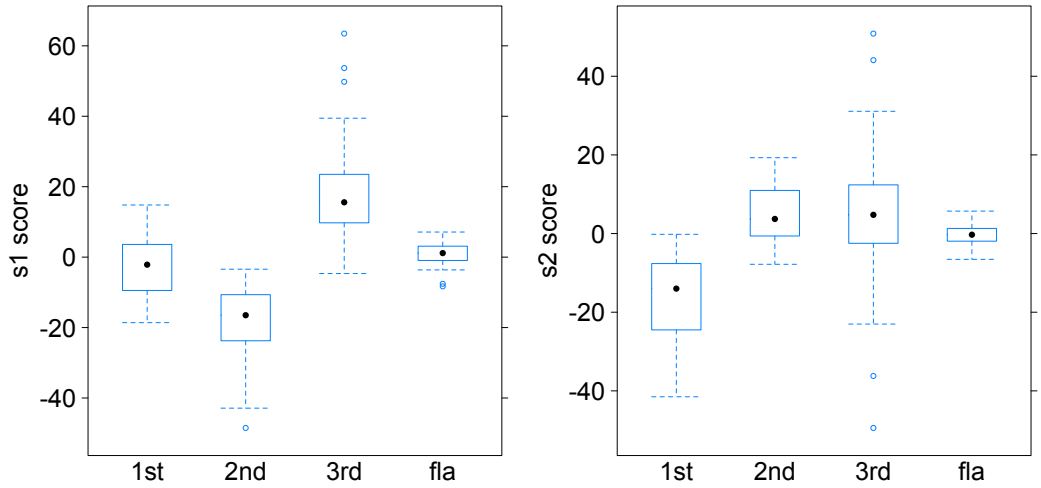


Figure 6.23 – s_1 and s_2 scores according to type

A Kruskal-Wallis rank sum test showed that s_1, s_2, and s_3 separate the curves above according their types:

- s_1: Kruskal-Wallis chi-squared = 138.1, df = 3, p-value < 2.2×10^{-16}
- s_2: Kruskal-Wallis chi-squared = 67.7, df = 3, p-value = 1.4×10^{-14}
- s_2: Kruskal-Wallis chi-squared = 10.4, df = 3, p-value = 0.015

A post-hoc analysis with the Wilcoxon test supports the classification of the curves according to the types that have been proposed in previous studies, as shown in Table 6.3.

Table 6.3: Pairwise comparisons of PC scores corresponding to *f0* curves in each group using Wilcoxon rank sum test (p-value adjustment: Bonferroni)

Types	**s_1**	**s_2**	**s_2**
1 & 2	<< 0.001	<< 0.001	< 0.05
1 & 3	<< 0.001	<< 0.001	0.2
2 & 3	<< 0.001	1	1
1 & F	0.7	<< 0.001	1
2 & F	<< 0.001	< 0.01	0.2
3 & F	<< 0.001	0.08	1

The table above shows that s_1 alone suffice to distinguish between the members of almost all of the pairs, except for Type 1 and Flat, which are distinguished by s_2. This constitutes compelling evidence that the classification of curves based on the plots shown in Figures 5.21 to 5.23 is indeed appropriate.

Figure 6.24 shows the mean curve $\mu(t)$ (upper left panel) along with the curves $PC_{1\text{-}3}(t)$ computed for the data under discussion in this section. As said earlier, the curves $PC_{1\text{-}3}(t)$ describe the main modes of variation in the sample and can be used to generate approximations for individual curves or models for groups of curves. Together, the curves $PC_{1\text{-}3}(t)$ account for 89% of the variability in the sample. Essentially like Figure 6.12, Figure 6.24 refers to both the three well-established classes of prosodic forms that function as TOP and to the Flat prosodic form. To be sure, there is much more variability in the sample of Figure 6.24 than in that of Figure 6.12, which represents only Type 1 and Type 2. This explains why the curve $\mu(t)$ in Figure 6.24 exhibits a more complex pattern.

Because a positive s_i score reinforces the shape of the corresponding curve $PC_i(t)$ and a negative s_i score inverts the shape of the corresponding curve $PC_i(t)$, by looking at Figure 6.24 we can guess the sign of s_i that a curve of a certain shape will have. For instance, Type 2 curves, which typically exhibit a rising movement, can be assumed to have negative s_1 values that invert $PC_1(t)$. Similarly, Type 1 curves can be assumed to have negative s_2 values that turn the fall-rise movement of $PC_2(t)$ into a rise-fall movement. The boxplots in Figure 6.23 above show that this seems to be actually the case.

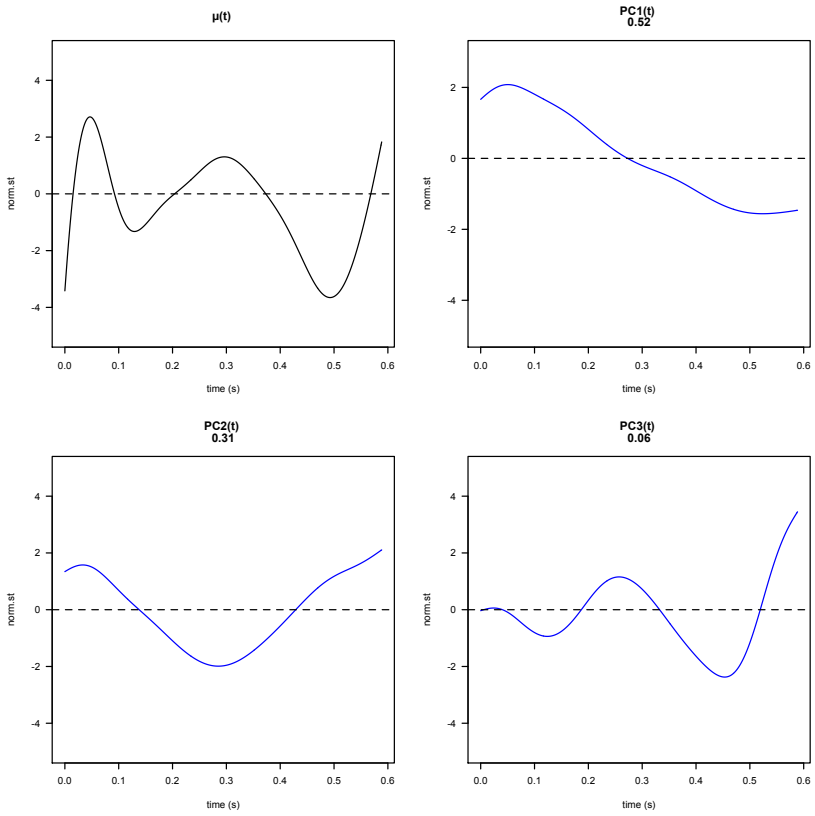


Figure 6.24 – $\mu(t)$, $PC_1(t)$, $PC_2(t)$, and $PC_3(t)$ computed for Type 1, Type 2, Type 3 and Flat, along with the proportion of variability accounted for by each curve $PC_i(t)$ (0.52, 0.31, and 0.06 respectively)

By plugging in eq. 1 – i.e. $f(t) = \mu(t) + s_1 \bullet PC_1(t) + s_2 \bullet PC_2(t)$ – the median s_1 and s_2 values for each class of TOP we obtain the models shown in Figure 6.25 below. Since $PC_3(t)$ only accounts for minor variations in the data, and s_3 are not required to distinguish the classes from one another, $PC_3(t)$ was not used for the generation of the models. The s_1 and s_2 values used for generating the curves in Figure 6.25, corresponding to the median s_i value of the curves of each class, are shown in the table below.

Table 6.4: Median s_i values used to generate the models in Figure 6.25

PC score	Type 1	Type 2	Type 3	Flat
s_1	-2.14	-16.5	15.6	1.14
s_2	-14	3.7	4.76	-0.3

By looking at Figure 6.25, we can see that the models and the diagrams presented in the beginning of this chapter closely correspond. It is worthwhile to mention that, even by using a large number of curves (183 in this case) and incorporating curves from different languages (see Table 6.2), as was done here, we arrive the forms that are described by the prior studies

presented in the beginning of this chapter (see RASO et al. 2017, for the latest published synthesis).

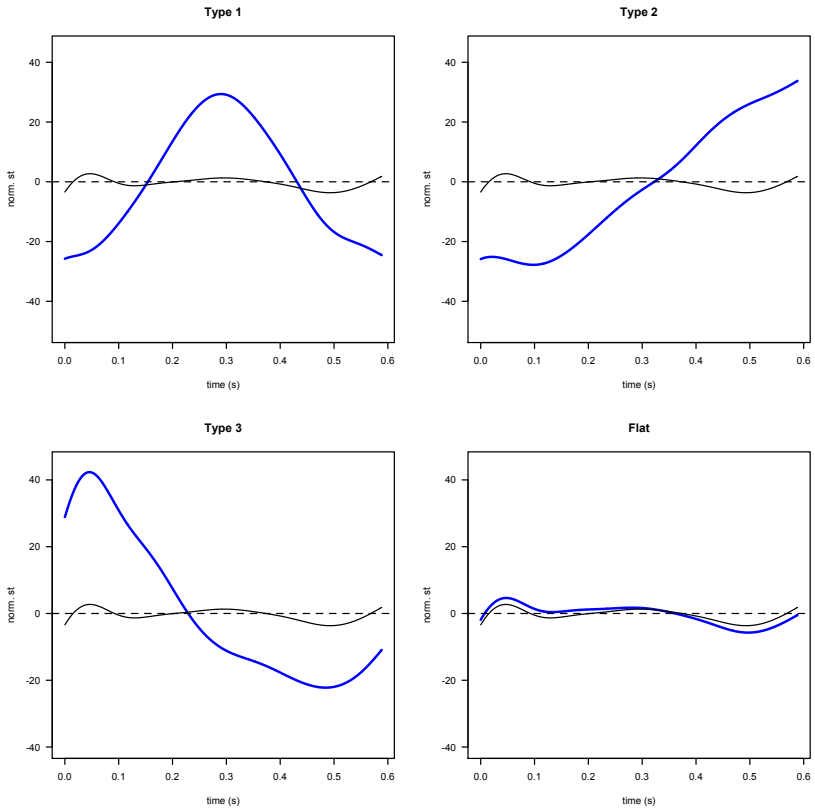


Figure 6.25 – Models for Type 1, 2, 3, and Flat. The mean curve $\mu(t)$ is plotted in all of the panels for comparison. The weights s_1 and s_2 used in each case correspond to the median corresponding score s_i computed for the curves of each type (see Table 6.4)

The analysis conducted so far shows that the group of Flat curves is similar to Type 3 in at least two respects: (i) both Flat and Type 3 have semi-nuclei that can be separated by a linking portion made up of non-nuclear syllables, and (ii) both classes exhibit an overall falling movement, differing basically in terms of movement range. If we look at the scatterplot in Figure 6.22, where s_1 and s_2 of all the curves are plotted, we see that the Flat curves occupy a region of the plot in which only a few Type 3 forms are seen. We take those similarities as indications that the Flat form actually constitutes a subtype within Type 3. We will return to this problem in the next chapter, in which we deal with the duration pattern of the syllables that constitute the prosodic units of TOP.

6.3.5 TOP units and Bound Comments

In this section, the prosodic units of TOP analyzed above are compared to prosodic units which exhibit an acoustic signal of continuity, but which convey illocutionary force – unlike TOP, which supply a cognitive domain for the interpretation of the illocution. Our aim is to counter the objection that the prosodic units of TOP may conceal different forms of continuity signals.

The prosodic units used for this comparison (see the bottom panel in Figure 6.26) correspond to bound comments (COBs), a type of illocutionary unit found in stanzas. A *stanza* is a terminated sequence that, as opposed to a regular utterance, is comprised of more than one illocution, at least one of which being a COB unit. The exact number of COBs in a stanza can vary substantially (see chapter 2) and they never constitute a melodic pattern together with other illocutionary units found in the same stanza.

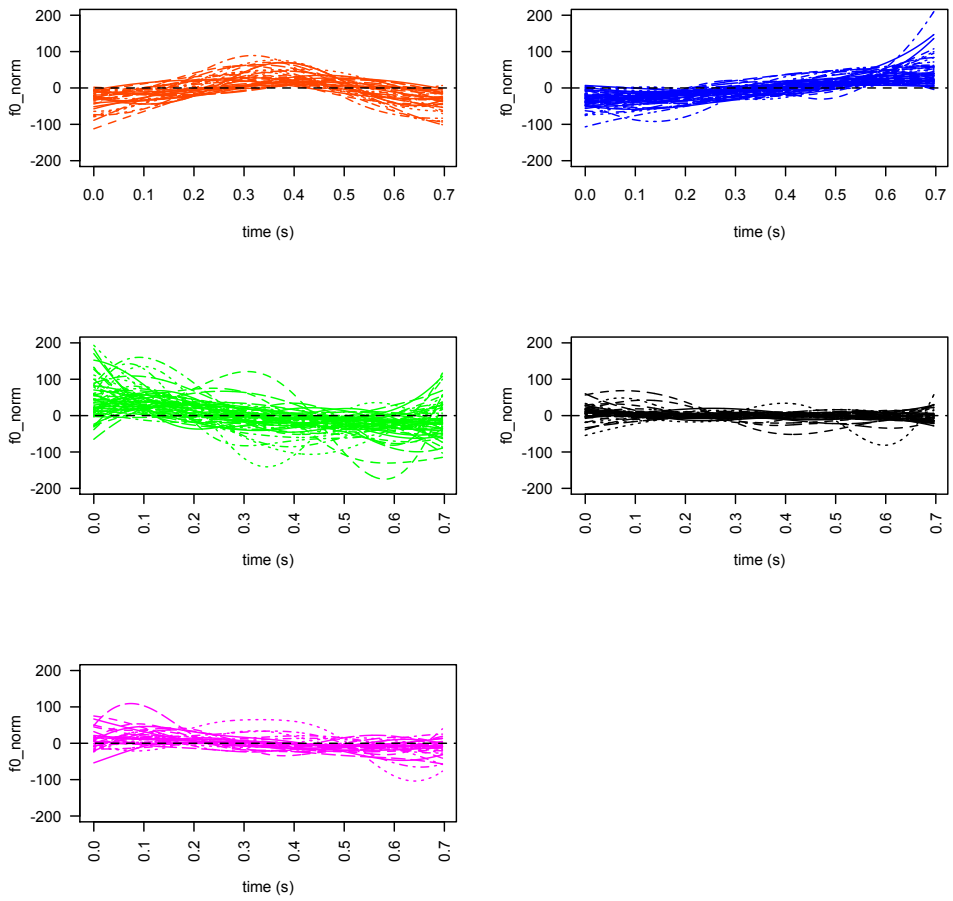


Figure 6.26 – Curve registration: Type 1 curves (top left panel), Type 2 curves (top right panel), Type 3 curves (middle left panel), Flat curves (middle right panel), and COB curves (bottom panel)

By subjecting 213 *f0* curves to FDA, we obtain the registered (smoothed and time-aligned) curves shown in Figure 6.26 above. The patterns that we see in the figure for Type 1, Type 2, Type 3, and Flat are the same as those shown in Figure 6.21 above. The panel at the bottom of the figure shows the *f0* curves of the COB units, the illocutionary units that feature the signal of continuity. We can notice that COBs resemble both Type 3 and Flat curves, in that they share the overall falling movement. The most relevant difference among these three classes is found in the *f0* range of each class: Type 3 has the widest range, Flat has the narrowest, whereas COB lies in between Type 3 and Flat.

We subjected these curves to F-PCA and found that the PC scores s_1 alone suffice to distinguish all the five classes. A Kruskal-Wallis rank sum test showed that both s_1 and s_2 distinguish the classes. A post-hoc analysis with the Wilcoxon provided further evidence for that. The results of the tests are shown below:

- s_1: Kruskal-Wallis chi-squared = 161.2, df = 4, p-value = 2.2e-16
- s_2: Kruskal-Wallis chi-squared = 75.0, df = 4, p-value = 2.08e-15
- s_3: type Kruskal-Wallis chi-squared = 14.0, df = 4, p-value = 0.008413

Table 6.5: Pairwise comparisons (classes of TOP curves and COB curves) using Wilcoxon rank sum test (p-value adjustment: Bonferroni)

	p-values		
Types	**s_1**	**s_2**	**s_3**
1 & 2	<< 0.001	<< 0.001	0.01
1 & 3	<< 0.001	<< 0.001	1
2 & 3	<< 0.001	1	1
1 & F	0.04	<< 0.001	1
2 & F	<< 0.001	< 0.01	< 0.01
3 & F	<< 0.001	1	1
1 & C	< 0.001	<< 0.001	1
2 & C	<< 0.001	0.07	1
3 & C	<< 0.01	1	1
F & C	0.04	1	1

Just as we did when only prosodic units of TOPs were under analysis, we generated models based on the curves $\mu(t)$ and $PC_i(t)$ obtained with F-PCA for TOPs and COBs. The curves $PC_{1\text{-}3}(t)$ representing both TOP and COB units account together for 92% of the

variability in the data (0.55, 0.32, and 0.05 respectively). Using the same technique as in the above section, we obtained the models shown in Figure 6.27 for all of the 5 classes of curves.

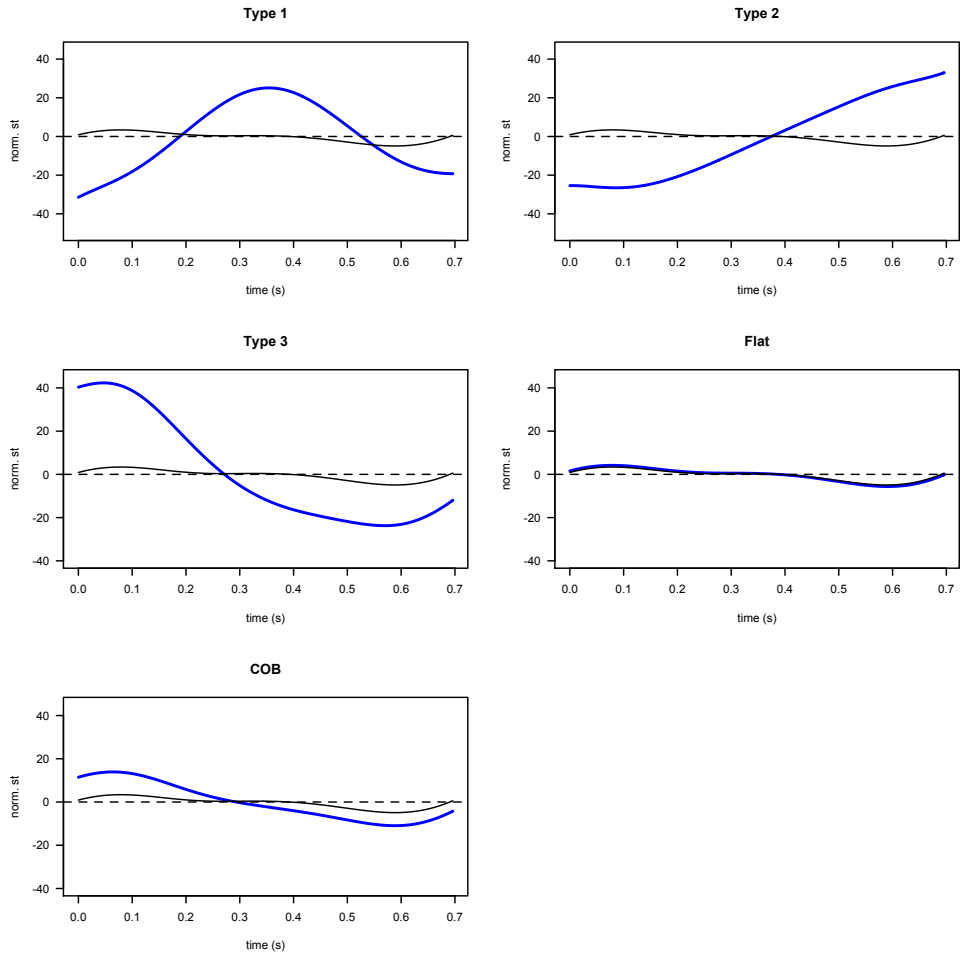


Figure 6.27: Models for Type 1, 2, 3, Flat, and COB. The mean curve $\mu(t)$ is plotted in each panel for comparison. The weights s_1 and s_2 used in each case correspond to the median score s_i computed for the curves of each group (see Table 6.6)

Table 6.6: Median s_i values used to generate the models in Figure 6.27

	Type 1	**Type 2**	**Type 3**	**Flat**	**COB**
s_1	6.25	18.41	-17.7	-0.46	-5.15
s_2	-13.53	4.66	4.65	-0.09	0.78

The comparison carried out in this section shows that TOP and COB are formally distinct information units. COB constitutes a unit whose nuclear syllables are inseparable from each other and always positioned at the end of the prosodic unit. It is similar in shape only to two classes of prosodic forms of TOP, namely Type 3 and Flat, whose nuclear syllables, on the contrary, are frequently structured as discontinuous semi-nuclei. This resemblance between

COB and TOP, however, is restricted to the overall shape of the curves, since their slopes are clearly distinct, as Figure 6.27 shows.

If it were true that TOPs merely consisted of different ways to mark continuity, then we would expect that COB curves fully overlapped with TOP curves, which does not occur, as can be seen in Figure 6.27. The region that registered curves of COB fill up in the plot does correspond to that filled up by certain curves of TOP, and precisely to the region occupied by the classes of TOP that exhibit two semi-nuclei (i.e. Type 3 and Flat). But since COBs have only one indivisible nucleus and accomplish an illocutionary function, they cannot be conflated with TOPs.

In the next chapter we analyze of the duration pattern of syllables, both nuclear and non-nuclear, that constitute the information unit of TOP. In addition to that, we tackle the problem involving the status of the Flat class taking into consideration the results of the study of syllable duration discussed therein.

7 The syllable duration facet of TOP

The concept of prosodic form, as discussed earlier (see chapters 2 and 6), encompasses not only specific patterns of *f0* movement but also the alignment of these patterns with the syllabic content of the unit as well as the longer perceived duration usually associated with nuclear syllables. This chapter deals with the relationship between syllable duration and the prosodic forms of TOP. Our goal is to verify whether the nuclear syllables of TOP are significantly longer than the non-nuclear ones, a characteristic that has been pointed out in many studies[49] but which, up until now, has not been statistically examined.

In addition, this study was designed as part of an attempt to better understand the Flat prosodic form, which will be done by comparing the duration of its syllables, both nuclear and non-nuclear, to the duration of Type 3 syllables, a prosodic form with which Flat shares similarities. Ultimately, we were interested in verifying whether Flat constitutes a Type 3 subtype or a prosodic form on its own right.

7.1 Materials and Methods

In this section we present the data, the statistical tools, and the procedures utilized in the duration analysis of the different prosodic forms that convey the function of TOP.

7.1.1 The dataset

When selecting the units for this part of the study, we had to be careful not to include in our sample cases of TOP featuring prosodic prominences associated with local emphasis outside the nucleus of the unit. This type of phenomenon is related to communication aspects with which we are not concerned here, and, by not being careful, we would be admitting into our sample a layer of complexity that would certainly hamper a proper evaluation of the duration of nuclear syllables, given that one of the prosodic parameters that signal local emphasis is precisely syllable lengthening.

In addition, *f0* variation and perceptual relevance usually interact (LEHISTE, 1975). Thus, considering that *f0* variation tends to affect perceived duration – with syllables exhibiting

[49] See Firenzuoli and Signorini (2003), Rocha (2012), Mittmann (2012), Cavalcante (2015), Raso et al. (2017).

dynamic *f0* patterns being perceived as longer –, we tried to keep to a minimum those TOPs showing too much *f0* variation outside the nucleus. Furthermore, we did not use monosyllabic TOPs for this analysis. It is worth mentioning that had little control over the units that would make up our sample, since we were dealing with naturally occurring data. Table 7.1 shows the distribution of the classes of prosodic forms in the sample according to language.

Table 7.1: Distribution of prosodic forms according to language and class

Language	Type 1	Type 2	Type 3	Flat	Total
AE	13	8	13	17	51
BP	11	11	12	3	37
EP	1	12	12	0	25
Total	25	31	37	20	113

Data from BP and EP were included in the sample due to the low frequency of Types 1 and 2 in the AE minicorpus, which did not allow for a balanced dataset. Also, we wanted to include the three instances of Flat found in the BP minicorpus. The 113 TOPs that were ultimately selected for the analysis added up to 690 syllables, and the number of syllables in each TOP ranged from 2 to 16 (M = 6.1, SD = 2.7); only 17 TOPs (i.e. 15% of the total) contained more than 8 syllables.

Finally, data from the IT minicorpus was not included, because we could not obtain a table containing the mean duration of phonetic segments in that language, an information that is required for the normalization that has to be carried out for comparing durations (see section 7.1.3).

7.1.2 The annotation

The data was segmented and annotated using Praat (BOERSMA; WEENINK, 2019). The annotation encompassed:

- the transcription of phonological syllables in ASCII
- indication of the boundaries and statuses of nuclear and non-nuclear syllables
- specification of the class of prosodic form of each nuclear syllable, i.e. Type 1 (“t1”), Type 2 (“t2”), Type 3 (“t31” for first semi-nucleus and “t32” for second semi-nucleus), or Flat (“tf1” for first semi-nucleus and “tf2” for second semi-nucleus)

- stress: monosyllabic word (“mono”), prestressed syllable (“pre”), stressed syllable (“str”), pos-stressed syllable (“pos”)
- classification of syllables as prominent (“s”) or non-prominent (“w”), which was particularly useful to distinguish stressed monosyllabic words from clitics
- the distance of a syllable from the boundary of the unit (in number of syllables).

In Figure 7.1, the first tier shows the transcription of the syllables in ASCII, the second tells whether a syllable is nuclear (“1”) or not (“0”), the third tells the class of the nuclear syllable (“t2”), the fourth tells that each syllable corresponds to a formal word, the fifth tells that the syllable/word [sens] ‘sense’ is prominent, and the sixth tier tells the distance from the boundary, where “0” means at the boundary. The tier at the bottom shows the orthographic transcription.

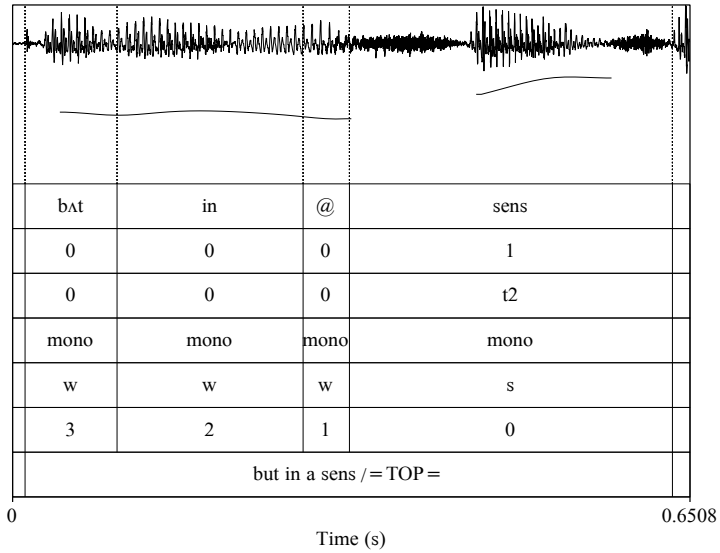


Figure 7.1 – TextGrid containing the annotation of a Type 2 TOP unit (afamdl01_62)

The reason for knowing the position of the syllables in relation to the boundary of the prosodic unit lies in the fact that, within a prosodic unit, the syllable sitting right at the boundary tends to be longer than syllables that sit farther away from it. Barbosa (2008) and Barbosa and Raso (2018), for example, point out that the speech rate decreases as prosodic units approach their boundary. This means that syllable lengthening may function also as a prosodic-break marker and thus cannot be considered solely a formal correlate of the information function of TOP.

7.1.3 Normalization

The normalization of the duration of the syllables analyzed herein was done using *SGDetector* (BARBOSA, 2006), a Praat script that uses the z-score normalization method (CAMPBELL, 1992; BARBOSA, 2006) to control for segments' intrinsic duration, thus rending their durations comparable to one another. The normalization of each syllable is obtained by means of the equation $z = \frac{dur - \sum_i \mu_i}{\sqrt{\sum_i var_i}}$, where *dur* is the syllable duration in milliseconds (ms), μ is the mean duration of all phones *i* in the syllable, and *var* is the variance of the phones. The descriptors μ and *var* come from Barbosa (2008) and Campbell (1992). As mentioned in the beginning in this chapter, data from Italian was not used in this analysis precisely because we do not have these descriptors for this language.

This script was slightly modified so as to enable it to deal with the variables with which it was not originally meant to deal, namely, *Nucleus*, *Nucleus Type*, *Stress*, *Prominence* and *Boundary*, which can be seen in Table 7.2 below, which shows the SGDetector output for the TOP unit shown in Figure 7.1.

Table 7.2: SGDetector output for the TOP unit shown in Figure 7.1

Syllable	Dur. (ms)	Dur. (z)	Nucleus	Nucleus type	Stress	Prominence	Boundary
b^t	88	-3,98	0	0	mono	w	3
in	179	-0,71	0	0	mono	w	2
@	45	-1,37	0	0	mono	w	1
sens	310	-2,43	1	t2	mono	s	0

In addition to raw and normalized durations – *Dur. (ms)* and *Dur. (z)* in the table above –, the script also provides the smoothed z-scores (not shown in the table). These, however, were not considered for this analysis. The main reason for this is that the smoothing technique used by the script (a five-period moving average) cannot be applied to prosodic units containing only a few syllables.

Finally, as the SGDetector was originally designed to be used with individual files, it was adjusted so that it could be applied to multiple files at once.

7.1.4 Statistical tools

The software used for the statistical analysis was R (R Core Team, 2019). We used a non-parametric analysis of variance (ANOVA), specifically by performing the Wilcoxon rank sum test, the Kruskal-Wallis rank sum test, the Pairwise post-hoc Wilcoxon signed-rank test and the Holm-Bonferroni method for p-value adjustment. The significance level established beforehand was 0.05.

7.2 Results

The ratio of nuclear to non-nuclear syllables in the sample is approximately 2:3, with nuclear syllables representing 40% of the total number of syllables that make up the sample (N = 690). This proportion, however, varies according to prosodic form type and language. Considering the languages separately, the proportions are as follows:

- AE: out of the 286 syllables, 124 (43%) are nuclear, 162 (47%) are not
- BP: out of the 251 syllables, 91 (36%) are nuclear, 160 (64%) are not
- EP: out of the 153 syllables, 63 (39%) are nuclear, 90 (61%) are not

If we consider the number of syllables grouped according to type of prosodic form, the counts are as follows:

- Type 1: out of the 147 syllables, 48 (33%) are nuclear, 99 (67%) are not
- Type 2: out of the 161 syllables, 53 (33%) are nuclear, 108 (67%) are not
- Type 3: out of the 286 syllables, 121 (42%) are nuclear, 165 (48%) are not
- Flat: out of the 96 syllables, 56 (58%) are nuclear, 40 (42%) are not

The higher proportion of nuclear syllables in Type 3 and Flat is due to the fact that these forms have two semi-nuclei, each of which usually composed of 2 syllables – and sometimes even more –, whereas the other prosodic form types have one indivisible nucleus that can be composed of one syllable alone, hence the equivalent proportions of nuclear and non-nuclear syllables for both Type 1 and Type 2.

Considering the sample as a whole, the mean duration of nuclear syllables (and standard deviation) is -0.3 (2.6) units of z-score, whereas the mean duration of non-nuclear syllables is -1.7 (1.4) units of z-score. This suggests that nuclear syllables do exhibit longer durations than

their non-nuclear counterparts, given that, when it comes to duration expressed in negative z-scores, the closer to zero, the longer the duration. A Wilcoxon rank sum test showed that this difference is statistically significant (see Table 7.3 below). This result corroborates the claim by previous studies (see e.g. RASO et al. 2017) that nuclear syllables of TOP units exhibit longer durations.

Table 7.3 below shows the mean normalized durations sorted by both nuclear status (i.e. nuclear or non-nuclear) and prosodic-form types.

Table 7.3: Averages of syllable durations (and standard deviations) in z-score units sorted by type of prosodic form. Type 3 and Flat semi-nuclei are considered together.

Form	**Nuclear syllables**	**Non-nuclear syllables**	**Wilcoxon test**
Type 1	0.6 (3.0)	-1.8 (1.3)	$W = 4122, p < 2.2x10^{-16}$
Type 2	1.3 (3.1)	-1.4 (1.5)	$W = 6289, p = 3.3x10^{-16}$
Type 3	-0.7 (2.1)	-1.6 (1.4)	$W = 12089, p < 2.2x10^{-16}$
Flat	-1.5 (1.5)	-2.4 (1.3)	$W = 432, p < 2.2x10^{-16}$
All	-0.3 (2.6)	-1.7 (1.4)	$W = 78560, p < 2.2x10^{-16}$

As seen in Table 7.3, the mean durations of nuclear syllables are consistently longer than the mean durations of non-nuclear syllables. The differences seen in the table are statistically significant, as the Wilcoxon-test column shows. Table 7.4 below shows the mean durations by language groups instead of prosodic-form type. The results are similar to those above, as the mean duration of nuclear syllables are also consistently and significantly longer than the mean duration of non-nuclear syllables.

Table 7.4: Mean durations of syllables (and standard deviations) in units of z-score sorted by nuclear status and language

Language	**Nuclear syllables**	**Non-nuclear syllables**	**Wilcoxon test**
AE	-1.8 (0.8)	-2.2 (1.0)	$W = 1258, p < 2.2x10^{-16}$
BP	1.4 (3.1)	-1.5 (2.0)	$W = 18398, p < 2.2x10^{-16}$
EP	0.4 (2.1)	-1.0 (1.2)	$W = 5625, p = 1.3x10^{-17}$

Let us now turn to the syllables of Type 3 and Flat semi-nuclei, whose durations are shown in Table 7.5 below. Regarding Type 3, pairwise comparisons using the Wilcoxon rank sum test showed that the duration of the syllables of the first semi-nucleus is not significantly different from the duration of non-nuclear syllables (p-value = 0.1), but it did show that the syllables of the second semi-nucleus are significantly longer in duration than non-nuclear

syllables (p-value = $2.7x10^{-7}$). So it seems reasonable to assume that the duration of second semi-nucleus is what accounts for the significant result related to the difference between Type 3 nuclear and non-nuclear syllables reported in Table 7.3 above. In addition, the test showed that there is a significant difference between the duration of Type 3 first and second semi-nuclei (p-value = 0.006).

Table 7.5: Mean syllable durations (and standard deviations) in units of z-score by semi-nuclei. Duration of non-nuclear syllables repeated for convenience

Type	**1st semi-nucleus**	**2nd semi-nucleus**	**Non-nuclear syllables**
Type 3	-1.3 (1.7)	-0.2 (2.3)	-1.6 (1.4)
Flat	-1.6 (1.5)	-1.3 (1.6)	-2.4 (1.3)

As for Flat, pairwise comparisons showed that there is no significant difference between the durations of the first semi-nucleus and the non-nuclear syllables (p-value = 0.1), no significant difference between the second semi-nucleus and the non-nuclear syllables (p-value = 0.07), and no significant difference between the first and the second semi-nuclei (p-value = 0.5). These non-significant results may be ascribed to sample size, particularly in the comparison between the syllables of the second semi-nucleus and the syllables that are non-nuclear, after all there are only 22 syllables of former group. For the sake of comparison, there are 34 syllables of the first semi-nucleus and 40 non-nuclear syllables.

We compared Type 3 semi-nuclei to Flat semi-nuclei, as well as the duration of these semi-nuclei to the nuclei of Type 1 and Type 2. Excluding all of the non-nuclear syllables, the results of the pairwise comparisons performed with the Wilcoxon rank sum test were as follows:

- Type 3 first semi-nucleus is shorter than
  - Type 1 nucleus (p-value = 0.04)
  - Type 2 nucleus (p-value = 0.0002)
- Type 3 second semi-nucleus is longer than
  - Flat first semi-nucleus (p-value = 0.003)*
- Flat first semi-nucleus is shorter than
  - Type 1 nucleus (p-value = 0.009)
  - Type 2 (p-value = 0.0002)
  - Type 3 second semi-nucleus of (p-value = 0.003)*
- Flat second semi-nucleus shorter than

- Type 2 (p-value = 0.02)

*The same information repeated for convenience

Based on the information that has just been shown, the duration of Flat second semi-nucleus is equivalent to all nuclear syllables, except for Type 2, which is the nucleus with the longest mean duration. The tendency in terms of the direction of the differences in each case can be seen in Figure 7.2 below.

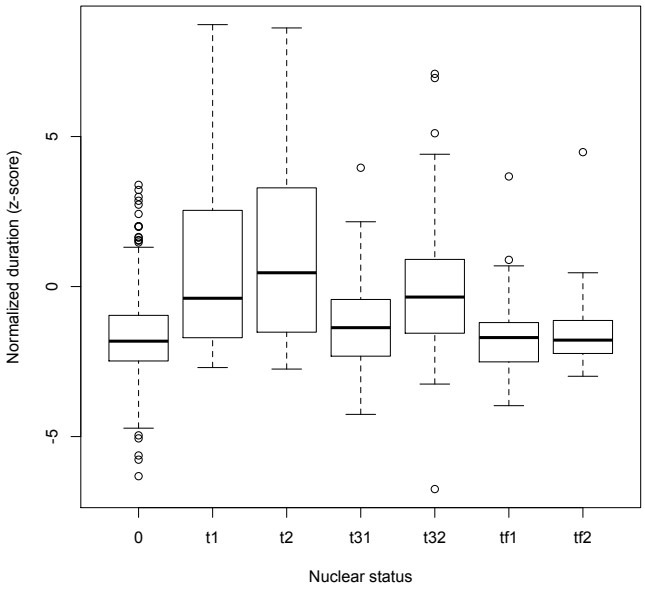


Figure 7.2 – Boxplots with syllable durations by prosodic-form type. Non-nuclear syllables represented by "0", Type 1 nucleus by "t1", Type 2 nucleus by "t2", Type 3 1st and 2nd semi-nuclei by "t31" and "t32", and Flat 1st and 2nd semi-nuclei by "f31" and "f32"

Considering that the nucleus of a TOP can be (and frequently is) made up of both stressed and unstressed syllables, we also did an analysis stress into consideration. We were particularly interested in examining the non-significant differences reported above between:

- Type 3 first semi-nucleus and non-nuclear syllables
- Flat first semi-nucleus and non-nuclear syllables
- Flat second semi-nucleus and non-nuclear syllables
- Flat first and second semi-nuclei

Before looking at these specific results, it must be said that, in our sample, stressed syllables are, unsurprisingly, significantly longer than unstressed syllables, with former having

a mean duration of -0.44 units of z-score and the latter having a mean of -1.5 (W = 79610, p-value < 2.2×10^{-16}).

The mean durations (and standard deviations) of stressed and unstressed syllables considered according to their information status are as follows:

a) stressed and nuclear syllables (n = 155): 0.2 (2.8)
b) stressed and non-nuclear syllables (n = 102): -1.5 (1.6)
c) unstressed and nuclear syllables (n = 123): -0.9 (2.2)
d) unstressed and non-nuclear syllables (n = 310): -1.7 (1.3)

Pairwise comparisons using the Wilcoxon rank sum test showed that the syllables in (a), which are both stressed and nuclear, are on average significantly longer than all other groups (p-value << 0.001 for all three pairs). The unstressed and nuclear syllables, in (c), are significantly longer than the syllables that are both unstressed and non-nuclear (p-value = 0.002), but they are not significantly longer than the syllables that are stressed and non-nuclear (p-value = 0.1). No statistically significant difference was found between the stressed non-nuclear syllables and the unstressed and non-nuclear syllables (p-value = 0.3).

By sorting Type 3 syllables by semi-nuclei and stress, and comparing the mean durations of each group thus formed using the Wilcoxon rank sum test, the only significant difference found was between stressed syllables of the second semi-nuclei and:

– unstressed non-nuclear syllables (p-value = 3.1×10^{-6})
– unstressed syllables of the first semi-nucleus (p-value = 0.0017)
– and stressed non-nuclear syllables (p-value = 0.0017)

The stressed syllables of the second semi-nuclei, with a mean duration of 0.44 (z-score), are longer than the syllables from all of the three other groups that have just been shown, whose mean durations are -1.7, -1.7, and -1.4 (z-score) (in the order they appear in the list immediately above). Type 3 first semi-nucleus and non-nuclear syllables – the two groups we were interested in examining given that their durations had not proven to be different in the computations carried out without taking stress into account – showed no significant difference between their durations.

By doing the same with Flat, we found no statistically significant difference in the pairwise comparisons. However, a Kruskal-Wallis rank sum test showed there may be significant differences in Flat grouped according to stress and informational status (chi-squared = 36.761, df = 5, p-value = 6.7×10^{-7}). The results of the pairwise comparison may be attributed to the small number of syllables in each group (e.g. Flat stressed nuclear syllables count only 4 occurrences). Given the limitations imposed by the size of our sample, we cannot arrive at reliable conclusions about the intricacies involving the duration pattern of Flat TOPs.

There is still one more variable to be considered, namely *boundary distance*, measured in number of syllables. In relation to the boundary of the prosodic units, the 278 nuclear syllables of our sample are distributed as follows: 100 of them sit right at the boundary, 90 of them sit one syllable away from the boundary, and the remaining 88 nuclear syllables sit more than two syllables away from the boundary.

Out of these 88 nuclear syllables found to be over two syllables distant from the boundary, 28 sit exactly two syllables away from the boundary and 24 sit three syllables away, which leaves 36 nuclear syllables at more distant positions and never forming groups containing than 9 cases. Table 7.6 displays these frequencies along with the equivalent information for non-nuclear syllables.

Table 7.6: Distribution of syllables in relation to prosodic boundary distance (percentages of columns totals)[50]

Position	**Nuclear**	**Non-nuclear**
at boundary	100 (36%)	12 (3%)
1 syll. away	90 (32%)	23 (6%)
2 syll. away	28 (10%)	83 (20%)
3 syll. away	24 (9%)	71 (17%)
over 3 syll. away	36 (13%)	223 (54%)
Total	278 (100%)	412 (100%)

As mentioned earlier, one of our concerns had to do with the effect of boundary proximity on the duration of the syllables, and, as seen in the above table, a significant proportion of nuclear syllables occur either at the prosodic boundary or very close to it. In trying to deal with this, we computed the difference between nuclear and non-nuclear syllables considering the positions. Table 7.7 shows these figures.

[50] Nuclear syllables included in the group "over 3 syll. away" never exceed 9 syllables within each position, the farthest from the boundary being 14 syllables distant.

Table 7.7: Mean syllable duration according to boundary distance

Group	Position	Nuclear	Non-nuclear
A	at boundary	0.2	-0.9
B	1 syll. away	0.3	-1.7
C	2 syll. away	-1.7	-1.6
D	3 syll. away	-1.5	-1.5
E	over 3 syll. away	-0.8	-1.8

A Kruskal-Wallis rank sum test indicated that the differences above are statistically significant (chi-squared = 83.022, df = 9, p-value = 4.05×10^{-14}), but the pairwise comparisons using the Wilcoxon test showed a significant difference only between nuclear and non-nuclear syllables in standing two syllables distant from the boundary (group B in Table 7.7). But if these syllables are considered together with the syllables at the boundary (i.e. if groups A and B are merged), then the pairwise comparisons show a significant difference between the durations of nuclear and non-nuclear syllables in this merged group (p-value = 0.008).

If groups A and B are put aside and all the other groups are merged, a statistically significant difference is observed in the comparison between nuclear and non-nuclear syllables (W = 22935, p-value < 2.2×10^{-16}). Furthermore, the pairwise comparisons of the durations of stressed nuclear syllables of groups A through E with one another show no statistically significant difference (p-value > 0.07 for all pairs). If the boundary proximity were the only aspect affecting the duration of these syllables, then a statistically significant difference between nuclear syllables closer to the boundary and those farther away would probably have been observed. Therefore, it appears that what is affecting the duration of the nuclear syllables that sit farther away from the boundary, making them similar in duration to the syllables that are nuclear and at the same time sit right at the boundary, is precisely their function as nucleus of the information unit in which they occur. We take that as evidence that syllable lengthening constitutes, in fact, one of the prosodic means by which the function of TOP is manifested.

In the next chapter, what has been discussed in this chapter about the duration facet of TOP will be integrated with the results from the analyses of the melodic shapes of the unit carried out in chapter 6 and with the agreement test presented in chapter 5.

8 Integrating the analyses

This chapter describes the information unit of TOP taking into consideration the results of the analyses discussed in the three previous chapters, therefore integrating the outcomes of the agreement test, the analysis of melodic shapes using FDA and F-PCA, and the analysis of syllable durations.

The issue of subjective interpretation is present in any kind of study that, like this dissertation, deals with abstract notions such as those corresponding to information structure categories. In order to address this problem, we conducted an inter-rater agreement test to assess the reliability of the detection of TOP in spontaneous speech, obtaining what we consider to be very good results, as suggests the kappa coefficient of 0.79 that expresses the overall agreement observed (see chapter 5).

It is worthwhile to mention that agreement tests are not very commonly found in the information structure literature, and the results of the relatively few that can be found tend to not suggest a suitable level of reliability (see LÜDELING et al. 2016 for a brief review). These tests, it should be added, are conducted, by and large, according to theoretical perspectives that are different from the perspective adopted in this study, which means, of course, that results should not be directly compared.

As shown in chapter 5, our agreement test was done with four raters, and it was based on 100 utterances from the three sections of the C-ORAL-BRASIL II corpus. Its results indicated substantial agreement, with kappa coefficients ranging from 0.66 to 0.80, depending on the section considered. The highest coefficient was computed for the 40 utterances from the Media section, and the lowest one for the 20 utterances from the Telephonic section. As for the 40 utterances from the Natural context section, its kappa coefficient was the same as that computed for all of the 100 utterances considered together (i.e. 0.79). We consider these results to be indicative of the empirical validity of TOP as defined by the Language into Act Theory (L-AcT).

When interpreting the agreement score relative to the annotation of the telephonic utterances (0.66), some aspects have to be taken into account. First, not only did the utterances from the Telephonic section were fewer in number than those from the other sections, but they also contained far fewer prosodic units (67 out of 541 prosodic units in the sample). In addition, the telephonic interactions from which these utterances were sampled were not particularly favorable to the occurrence of TOP, given that the frequency of this information unit is

associated with exchanges that are complex in terms of textual construction, low in interactivity, and less context dependent. The reason why it is important to attend to those aspects is that the kappa statistic is susceptible to the prevalence of a phenomenon in such a way that, for rare findings, a small kappa coefficient does not necessarily reflect a low degree of agreement (VIERA; GARRETT, 2005). Therefore, the kappa coefficient computed for the agreement in the telephonic utterances (0.66) has to be interpreted considering the low prevalence of TOPs in the interactions from which these utterances were sampled. Indeed, no annotator found more than two TOPs in these utterances. It is worth mentioning, however, that according to usual standards (see LANDIS; KOCH, 1977), 0.66 still falls within the range of kappa coefficients considered to represent substantial agreement.

The analysis of the melodic forms was conducted using FDA, F-PCA and ANOVA, with no information regarding the prior classification of the curves in the sample. The goal was to assess the descriptions made in previous studies regarding the melodic correlates of the function of TOP, which had not been validated statistically (see RASO et al., 2017). We looked at spontaneous speech data from corpora of four different language, namely American English, European Portuguese, Brazilian Portuguese, and Italian.

The time-aligned and smoothed representations of the input curves obtained with FDA (see Figure 6.23) showed that the curves associated with each prosodic form exhibit common patterns of *f0* movements. Furthermore, the patterns revealed by FDA closely resemble those described in previous studies, that is, a rise-fall movement for Type 1, a rising movement for Type 2, and, for Type 3, a first semi-nucleus characterized by high *f0* values and a second semi-nucleus with low *f0* values. As for Flat, FDA shows the curves of this group clustering together around the mean *f0* curve, exhibiting the first semi-nucleus with slightly higher values than the second one, thus suggesting a pattern somewhat similar to that of Type 3 but with a less steep *f0* slope.

One of the questions that we have addressed has to do with the status of Flat as a form in its own right or as subgroup within Type 3. Although we have not been able to come to a final conclusion, this study does give us insight into Flat TOP units and their status. First, the PC scores s_1, s_2, and s_3 obtained using F-PCA allowed for us to establish, by means of pairwise comparisons using the Wilcoxon rank sum tests, that the curves associated with each class of prosodic form (Flat included) are separable from the curves associated with the other classes of forms. The separability is warranted by s_1 and s_2, which together account for 83% of the

variability observed in the analyzed data. Based on this analysis, Flat does constitute an independent group, as it is significantly different from all of the other classes (see chapter 6).

In order to counter the objection that the forms of TOP may conceal different forms of acoustic signals of continuity, we introduced another set of curves to the analysis. These curves are associated with what is known within L-AcT as bound comments (COB), a special sort of illocutionary unit that occurs in stanzas and features a continuity signal. We then repeated both FDA and F-PCA, generating models based on the median PC score value of the curves associated with each class. The patterns obtained with the addition of COBs to the data are shown in Figure 6.24, which repeated below for convenience as Figure 8.1.

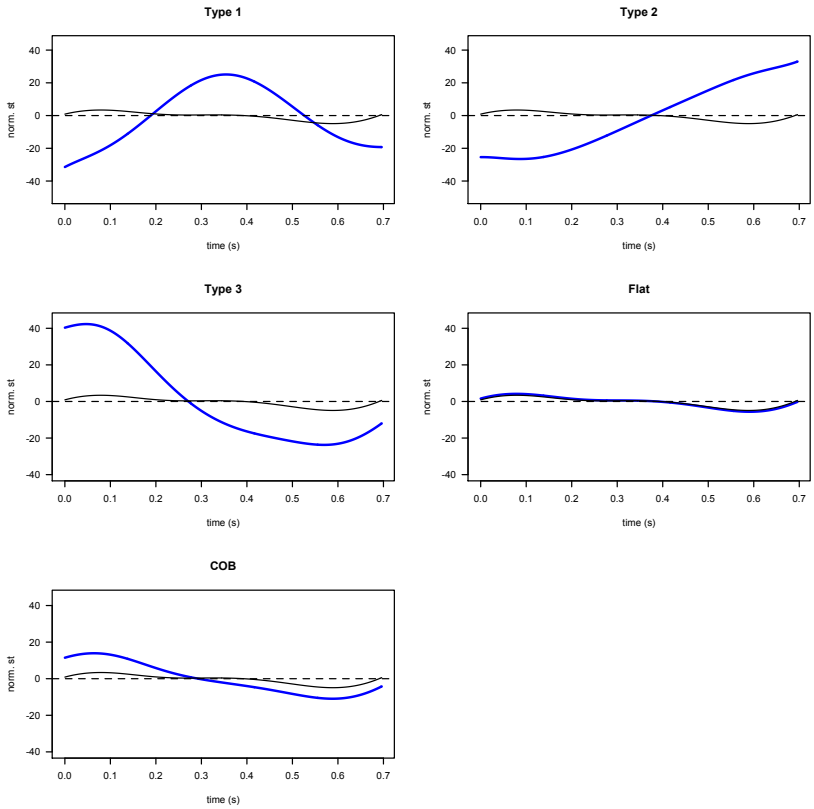


Figure 8.1 – Model curves for Type 1, 2, 3, Flat, and COB. The mean curve $\mu(t)$ is repeated in every panel. The weights s_1 and s_2 used in each case correspond to the median PC score value computed for the curves of each class (see Table 6.6 in chapter 6 for details).

The analysis of variance and the post-hoc pairwise comparisons showed that the curves grouped by class continue to be separable just as they were prior to the addition of COB curves. By looking at the figure above, we can see that the model obtained for COBs is clearly different from Type 1 and Type 2 models, but in some ways, it resembles both to Type 3 and Flat.

The resemblance in melodic shape, however, does not mean that the forms are equivalent, given that both Type 3 and Flat have two often discontinuous semi-nuclei, whereas COB has only one nucleus, which is indivisible, we must stress. In addition to that, the ranges of *f0* values associated with Type 3, Flat, and COB differ from one another, which is reflected in their significantly different PC scores. Therefore, we conclude that TOP cannot be confounded with COB, as they are different both in terms of function and prosodic from. Furthermore, it is remarkable that TOP and COB are distinguishable on the basis of phonetic properties with no need for semantic or pragmatic concepts.

All that has been said concerning TOP and COB does not mean that TOP has no continuity signal. On the contrary, it indicates that, besides the continuity signal, its form accomplishes a function that is different from the function performed by another category (i.e. COB), whose form, too, includes a continuity signal.

In chapter 7, we dealt with the durational facet of TOP. We were interested in examining the allegedly longer duration associated with the nuclear syllables of the unit, a perceptually salient feature which has been noted in different studies, but which had not been examined statistically before.

We analyzed a dataset containing TOPs from AE, BP, and EP. No IT data was examined, both because the IT minicorpus was not fully completed at the time of analysis and also because we do not have the required mean and the standard deviation values relative to IT phones, which are required for normalizing syllable durations and thus enabling their comparison.

The results of the comparison of syllable durations showed that, on average, nuclear syllables are significantly longer than non-nuclear ones. Considering the nuclear and non-nuclear syllables of each class of prosodic form separately, the results were similar, as seen in Table 7.3, repeated below as Table 8.1 for the sake of convenience.

Table 8.1: Mean normalized durations (and standard deviations) of syllables according to nucleus and type. Type 3 and Flat are not distiguished by semi-nuclei.

Form	**Nuclear syll.**	**Non-nuclear syll.**	**Wilcoxon test**
Type 1	0.6 (3.0)	-1.8 (1.3)	$W = 4122$, $p < 2.2 \times 10^{-16}$
Type 2	1.3 (3.1)	-1.4 (1.5)	$W = 6289$, $p = 3.3 \times 10^{-16}$
Type 3	-0.7 (2.1)	-1.6 (1.4)	$W = 12089$, $p < 2.2 \times 10^{-16}$
Flat	-1.5 (1.5)	-2.4 (1.3)	$W = 432$, $p < 2.2 \times 10^{-16}$
All	-0.3 (2.6)	-1.7 (1.4)	$W = 78560$, $p < 2.2 \times 10^{-16}$

Considering Type 3 syllables grouped according to semi-nucleus, the pairwise comparisons showed that the syllables of the second semi-nucleus are significantly longer than

both the syllables of the first semi-nucleus (p-value = 0.006) and the non-nuclear syllables (p-value = $2.7x10^{-7}$). No statistically significant difference between the first semi-nucleus and the non-nuclear syllables was observed (p-value = 0.1). This may be taken as an indication that the set of prosodic parameters that defines the first semi-nucleus is a little different from the set that defines the second semi-nucleus. This does not appear to constitute an actual problem, especially if we take into account that the first semi-nucleus is more distant from the unit's boundary, and that being close to the boundary typically correlates with having a comparatively longer duration.

Regarding the duration of Flat, the pairwise comparisons of the first and second semi-nuclei and non-nuclear syllables showed no statistically significant difference. That is, despite the fact that Flat nuclear syllables – without distinguishing semi-nuclei – tested significantly longer than their non-nuclear counterparts, once they were separated according to semi-nucleus and then compared to the non-nuclear syllables, the differences between the groups taken two at a time did not prove to be statistically significant (p-values ≥ 0.07).

The comparison for which *p-value* = 0.07, i.e. that which is closer to the significance level of 5%, is precisely between the second semi-nucleus and non-nuclear syllables. This is unsurprising, after all the second semi-nucleus is closer to the boundary and the first semi-nucleus, while sitting farther away from the boundary, is still under the effect of the function of TOP which it helps to accomplish. However, this result may be partly due to the small number of syllables in each of these groups (i.e. Flat two semi-nuclei and non-nuclear syllables), particularly in the second semi-nucleus group[51] (see WASSERSTEIN; LAZAR, 2016). It is important to note that Flat is a very uncommon form, appearing to be almost absent in some languages, with obvious consequences for our sample. In fact, it was only after Flat was observed in AE, where it corresponds to 9% of the curves observed, that a few occurrences were observed in the other languages.

Considering all classes and disregarding all of their non-nuclear syllables, thus focusing only on syllables that are either semi-nuclei or nuclei, the Flat second semi-nucleus only tested significantly different when compared to syllables of the Type 2 nucleus, which is precisely the group containing featuring the longest mean duration. This suggests that it is reasonable to believe that the duration of Flat nuclear syllables may actually be comparable to the nuclear

[51] The second semi-nucleus group contains only 22 syllables, while the first semi-nucleus group and the non-nuclear syllable group have 34 and 40 syllables, respectively.

syllables of other forms, and that the non-significant results computed may turn out differently in future analyses based on a sample featuring a greater number of curves.

If in the future a larger sample containing more Flat curves ends up showing to be true that the first semi-nucleus does not exhibit a significant difference in relation to the non-nuclear syllables and the second semi-nucleus proves to be significantly longer than both the first semi-nucleus and the non-nuclear syllables, we will probably be able to conclude that Flat has a durational pattern akin to Type 3. In that case, we would have two forms exhibiting a common duration pattern and melodic curves that, although separable on the basis of PC scores, still have fairly similar shapes.

In trying to tackle the effect of boundary proximity, we did the tests disregarding the syllables located close to the boundary and considering only those syllables sitting two or more syllables away from the boundary. The comparison between syllables meeting this condition – disregarding the grouping according to class – showed that nuclear syllables are in fact longer than non-nuclear syllables ($W = 22935$, p-value $< 2.2e\text{-}16$).

The comparison of the duration of stressed nuclear syllables grouped according to boundary proximity showed no significant difference between the stressed nuclear syllables that are closer to the boundary and those that are farther away. We take it as evidence for the presence of an effect other than that of boundary proximity, after all, if there were none, then we would expect some significant difference between the groups of syllables closer to the boundary and those farther away. Naturally, we assume that what is causing this effect is that these syllables are accomplishing the function of TOP. In addition, grouping the syllables according to stress (stressed or unstressed) and nucleus (nuclear or non-nuclear), unstressed nuclear syllables are significantly longer than their non-nuclear counterparts (p-value = 0.002), which appears to constitute yet more evidence for there being an effect associated with the function of TOP.

The analysis of Flat syllables grouped according to semi-nuclei and stress showed a significant difference in the Kruskal-Wallis test (chi-squared = 36.761, df = 5, p-value = $6.7\ 10^{-7}$) but no significant difference in the pairwise comparisons. As in the pairwise comparison of Flat syllables without considering the variable stress, the non-significance of the results may be ascribed to insufficiency of data, given that Flat stressed nuclear syllables count only 4 occurrences.

In summary, the analyses carried out in this dissertation contemplated the two main formal facets of TOP: melodic shape and syllable duration. Furthermore, we conducted an inter-

rater agreement test in which we obtained very good kappa coefficients, showing that TOP does have an objective reality.

Based on the evidence we found, we can conclude that there are at least three classes of prosodic forms of TOP, namely Type 1, Type 2, and Type 3. As for Flat, we could neither establish its status as a form in its own right nor determine it constitutes a sub-form of Type 3. The analysis of variance of PC scores suggests it is a separate group, but the shape of its curves and, to a certain extent, the durational pattern of the first semi-nucleus suggest a strong similarity with Type 3.

Another prosodic parameter described as relevant for TOP is intensity. We were not able to look into it in the present study. This prosodic parameter may be of particular importance to further refine the description of Flat. Also it could help us refine the description of Type 3, particularly its first semi-nucleus, which did not show a statistically significant difference from the duration of the non-nuclear syllables.

One of the reasons to believe in the relevance of intensity for the description of TOP is that there seems to be a positive association between intensity and perceived pitch (HOWARD; ANGUS, 2017, p.153)[52]. Thus, the perceptually noticeable pitch variation in units that exhibit practically no *f0* variation (i.e. the Flat curves) may be accounted for by the interplay between intensity and *f0*.

It would be interesting to be able to replicate this study by looking not only at more TOP units but also at TOP units from languages other than those examined here. Fortunately, in addition to the new version of the IT minicorpus, a revised version of the Spanish minicorpus is underway. Thus, we will soon be able to sample TOPs from the entire IT minicorpus – which may add a few Flat forms to the sample – and from the Spanish minicorpus. Apart from increasing the likelihood of finding more Flat forms, the analysis of other languages may shed light on aspects of TOP which are simply absent or too subtle in the languages so far examined.

We have learned that the prosodic forms of TOP constitute intricate objects. The study of the unit within L-AcT first began by on the basis of IT data and, as data from more languages have been examined, more complexity has been revealed. This complexity encompasses both the varying frequency of types in each language as well as the interplay of prosodic parameters and even the existence of different prosodic forms.

---

[52] We thank Donna Erickson for pointing to this reference.

Finally, it is hypothesized that the frequency of prosodic forms in a language may be a function of the rhythmic structure the language. This hypothesis has been raised based on the observation that Type 1 is the most frequent form in IT (55%), Type 2 in BP (47%), and Type 3 is the most frequent form in EP (58%) and especially in AE (72%). In light of that, it appears to be the case that the more syllable-timed a language is, the more frequent Type 1 is, and the more stressed-timed a language is, the more frequent Type 3 is. This hypothesis should probably be considered in future research.

9 Final words

In this dissertation, we looked into the information unit of TOP, as defined by the Language into Act Theory (L-AcT), by conducting statistical analyses using data from American English, Italian, Brazilian Portuguese, and European Portuguese. The data were extracted from spontaneous speech corpora compiled according to the theoretical tenets and methodological standards of L-AcT. In addition, we compared the approach of L-AcT to the study of information structure to some of the other major approaches found in the literature.

Chapter 1 presented a critical review of the information structure literature, thus providing a backdrop for the assessment of the theoretical perspective adopted in this study. We discussed some of the most recurrent terms and concepts in the literature, foreshadowing some of the ways in which L-AcT stands out as an empirically grounded and theoretically coherent framework.

Chapter 2 provided a comprehensive description of the L-AcT framework. Two of the most relevant aspects of the framework are the recognition of the illocution as the central element of speech and the importance ascribed to prosody, which demarcates prosodic units, signals the illocution and participates in the signaling of other information functions. The chapter provided a first description of TOP according to L-AcT, which defines it as the information unit that supplies a cognitive domain for the interpretation of the illocution and that is realized by prosodic forms characterized by specific prosodic features.

Chapter 3 discussed some the different notions that have been associated with the term topic, showing the importance of the focus given by L-AcT on illocution and prosody for a definition of TOP in terms of pragmatic relationship of *aboutness*, that is, a relationship between a cognitive domain and an illocution.

Chapter 4 presented the corpora used in this study, which are members of the C-ORAL family of corpora and document spontaneous speech in various communicative situations. All of them feature prosodic segmentation into utterances and prosodic units and text-to-speech alignment on the utterance level. The minicorpora, in particular, also feature annotation of information functions. In addition, we discussed the revision that made in the AE and BP minicorpora prior to the retrieval of information units for our analyses and briefly described the old and new versions of the IT minicorpus, which was used in the analysis of melodic forms reported in chapter 6.

In chapter 5, we presented the results of the inter-rater agreement test conducted in order to assess the reliability of the detection of TOP in spontaneous speech. The test was based on 100 utterances from the three sections of the C-ORAL-BRASIL II. Our results showed substantial agreement, according to usual standards, among the four annotators who took part in the test, with kappa coefficients ranging from $k = 0.66$ to $k = 0.8$, which varied according to the section from which the utterances were obtained. The section with the highest agreement level was *Media*, and the one with the lowest level was *Telephonic*. The kappa coefficients computed for the *Natural context* section and the three sections considered together are the same, specifically $k = 0.79$. We argued that the lower coefficient in the case telephonic utterances should be interpreted taking into consideration the kappa statistic sensitivity to low-prevalence phenomena, as a small kappa coefficient does not necessarily reflect a low degree of agreement for rare findings.

In chapter 6, we presented the analysis of the prosodic forms of TOP, for which we used Functional Data Analysis (FDA), Functional Principal Components Analysis (F-PCA), and Analysis of Variance (ANOVA) to examine the melodic correlates of the function of TOP. This study presents the first statistical validation and a synthesis of the descriptions made in previous studies regarding the melodic patterns associated with TOP. A valuable aspect of the analysis reported in chapter 6 is that the statistical techniques that were used requires no *a priori* assumption as to the exact portions of the f0 curves that are functionally relevant. Also, the computations are made without any information regarding the classes that the curves supposedly belong to. Class-membership information is introduced only in order to check whether the patterns revealed by the techniques actually correspond to the classification conducted beforehand.

In chapter 7, we dealt with the durational facet of TOP. We were interested in looking into the allegedly longer duration associated with the nuclear syllables of TOP, a perceptually plausible feature which has been noted in different studies but had not been examined statistically before. Our results showed that, on average, nuclear syllables are indeed significantly longer than non-nuclear syllables. Considering each class of prosodic form separately, the results are similar, and the nuclear syllables within each type are significantly longer than their non-nuclear counterparts.

The results obtained with the analyses discussed in chapters 5 and 6 allowed us to conclude that there are at least three classes of prosodic forms of TOP, namely Type 1, Type 2, and Type 3. The analysis of PC scores using ANOVA suggests that Flat constitutes a group

separate from Type 3. However, the overall melodic shape of Flat curves and the durational pattern of their first semi-nucleus suggest the opposite.

In the future, we hope to be able to arrive at a final answer to the issues involving Flat, possibly by examining the parameter of intensity, which was not done in this work. As previously said, there seems to be a positive association between intensity and perceived pitch, and this may explain why we perceive pitch variation in Flat units.

The prosodic forms of TOP, as this dissertation has shown, constitute a complex phenomenon. We believe that future research should not only focus on intensity but also try to enrich the data both in terms of languages represented and number of units. This is particularly important if we wish to refine our understanding of Flat, as it is far from being the most frequent of prosodic forms. Examining data from different languages may also bring new insights into TOP, particularly into the prosodic features of the unit.

References

ADRIAANS, Pieter; BENTHEM, J. F. A. K. van. Philosophy of information. *Handbook of the Philosophy of Science* 8, 2008.

ADRIAANS, Pieter, Information. *The Stanford Encyclopedia of Philosophy.* ZALTA, E. N. (ed.), 2017. <https://plato.stanford.edu/archives/fall2017/entries/information/>.

ARISTOTLE. Categories. In: BARNES, Jonathan (ed.). *The Complete Works of Aristotle*. Princeton: Princeton University Press, 1984, p. 22–70.

ARNOLD, J. E.; KAISER, E.; KAHN, J. M.; & KIM, L. K. Information structure: linguistic, cognitive, and processing approaches. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(4), 2013, p. 403-413.

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

BACELAR DO NASCIMENTO, F. et al. The Portuguese Corpus. In: CRESTI, EMANUELA; MONEGLIA, MASSIMO (Org.). *C-ORAL-ROM: integrated reference corpora for Spoken Romance Languages*. Amsterdam / Philadelphia: John Benjamins, 2005. p. 163–208.

BARBOSA, Plínio A. *Incursões em torno do ritmo da fala*. Campinas: Pontes, 2006.

BARBOSA, Plínio. A. Prominence-and boundary-related acoustic correlations in Brazilian Portuguese read and spontaneous speech. In *Proc. Speech Prosody*, 2008, p. 257-260.

BARBOSA, Plínio. A; RASO, T. Spontaneous Speech Segmentation: Functional and Prosodic Aspects With Applications for Automatic Segmentation/A segmentação da fala espontânea: aspectos prosódicos, funcionais e aplicações para a tecnologia. *Revista de Estudos da Linguagem*, 26(4), 2018, p. 1361-1396.

BARNES, Jonathan (ed.). *The Complete Works of Aristotle*. Princeton: Princeton University Press, 1984.

BIBER, Douglas; JOHANSSON, Stig; LEECH, Geoffrey; CONRAD, Susan; FINEGAN, Edward. *Longman Grammar of Spoken and Written English*. Harlow: Longman, 1999.

BICK, Eckhard. A anotação grammatical do C-ORAL-BRASIL In: RASO, Tommaso; MELLO, Heliana (eds.). *C-ORAL-BRASIL I: corpus de referência de português brasileiro falado informal*. Belo Horizonte: UFMG, 2012, p. 223-254.

BICK, Eckhard. The grammatical annotation of speech corpora: techniques and terspectives. In: RASO, Tommaso; MELLO, Heliana (eds.). *Spoken corpora and linguistic studies*. John Benjamins Publishing Company, 2014, p. 106-128.

BOERSMA, P.; WEENINK, D. *Praat: doing phonetics by computer: computer program*. Version 5.4.21 Amsterdam: University of Amsterdam, 2019. Available at: http:// www.praat.org/.

BOLINGER, Dwight L. English prosodic stress and Spanish sentence order. *Hispania*, 1954, p. 152–156.

BRENTANO, Franz Clemens. *Psychologie vom empirischen Standpunkt*. Vol. 1. Berlin: Duncker & Humblot, 1874.

CAMPBELL, Wilhelm Nick. *Multi-level timing in speech.* PhD dissertation. University of Sussex, 1992.

CARUANA, Fausto; BORGHI, Anna. *Il cervello in azione: introduzione alle nuove scienze della mente*. Bologna: Il Mulino, 2016.

CAVALCANTE, Frederico A. *The topic unit in spontaneous American English: a corpus-based study*. 2015. 184f. M.A. thesis. Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2015.

CAVALCANTE, Frederico A.; RAMOS, Adriana Couto. The American English spontaneous speech minicorpus: architecture and comparability. *CHIMERA: Romance Corpora and Linguistic Studies*, 3(2), 2016, p. 99–124.

CHAFE, Wallace L. Language and consciousness. *Language*, 1974, p. 111–133.

CHAFE, Wallace L. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In: LI, Charles N. (Ed.). *Subject and topic*. New York: Academic Press, 1976.

CHAFE, Wallace L. *Discourse, consciousness, and time: the flow and displacement of conscious experience in speaking and writing*. Chicago IL: The University of Chicago Press, 1994.

CHOMSKY, Noam. Deep structure, surface structure, and semantic interpretation. In: STEINBERG, DANNY; JAKOBOVITS, LEON (org.). *An interdisciplinary reader in Philosophy, Linguistics, and Psychology*. Cambridge: Cambridge University Press, 1971. p. 183–216.

CHOMSKY, Noam. *Studies on Semantics in Generative Grammar*. Paris: Mouton, 1972.

CHOMSKY, Noam. *Knowledge of language: Its nature, origin, and use*. Westport, Connecticut/London: Praeger, 1986.

COHEN, Jacob. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), p. 3746, 1960.

CRESTI, Emanuela. *Corpus di italiano parlato*. Firenze: Accademia della Crusca, 2000.

CRESTI, Emanuela. Notes on lexical strategy, structural strategies and surface clause indexes in the C-ORAL-ROM spoken corpora. In: CRESTI, E.; MONEGLIA, M. (eds.). C-ORAL-ROM. *Integrated reference corpora for spoken Romance languages*. Amsterdam: John Benjamins, 2005, p. 209–256.

CRESTI, Emanuela. The definition of focus in Language into Act Theory (LAcT). In: MELLO, HELIANA; PANUNZI, ALESSANDRO; RASO, TOMMASO (Org.). *Pragmatics and Prosody: Illocution Modality, Attitude, Information Patterning and Speech Annotation*. Firenze: Firenze University Press, 2011.

CRESTI, Emanuela. Syntactic properties of spontaneous speech in the Language into Act Theory: data on Italian complements and relative clauses. In: RASO, T.; MELLO, H (Eds.). *Spoken corpora and linguistic studies*. Amsterdam: John Benjamins, 2014, p. 365-410.

CRESTI, Emanuela; GRAMIGNI, Paola. Per una linguistica corpus based dell'italiano parlato: le unità di riferimento. In: F. Albano Leoni; F. Cutugno; F. Pettorino; M. (acd) Savy (Eds.). *Il parlato italiano*. Atti del Convegno internazionale Napoli, 13-15 febbraio 2003, CD-ROM. Atti... Napoli: D'Auria, 2004, p.1–23.

CRESTI, Emanuela; MONEGLIA, Massimo (Eds.). C-ORAL-ROM. *Integrated reference corpora for spoken Romance languages*. Amsterdam: John Benjamins, 2005.

CRESTI, Emanuela; MONEGLIA, Massimo. Informational patterning theory and the corpus-based description of spoken language: The compositionality issue in the topic-comment pattern. In: PANUNZI, Alessandro; MONEGLIA, Massimo. *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*. Firenze: Firenze University Press, 2010, p. 13-45.

CRESTI, Emanuela; RASO, Tommaso. *Annotation of information units through IPIC*. 2012. http://lablita.dit.unifi.it/ipic/.

DAHL, Östen. How do noun phrases refer? *Logical Grammar Reports*, v. 21, 1976a.

DAHL, Östen. What is new information? *Reports in Text Linguistics*, v. 72, 1976b.

DOMANESCHI, Filippo. *Presuppositions and Cognitive Processes: Understanding the Information Taken for Granted*. London: Macmillan, 2016.

DU BOIS, John W.; CHAFE, Wallace L.; MEYER, Charles; THOMPSON, Sandra A.; ENGLEBRETSON, Robert; MARTEY, Nii. *Santa Barbara corpus of spoken American English, Parts 1-4*. Philadelphia: Linguistic Data Consortium, 2000-2005.

ERTESCHIK-SHIR, Nomi. *The dynamics of focus structure*. Cambridge: Cambridge University Press, 1997.

ERTESCHIK-SHIR, Nomi. *Information structure: The syntax-discourse interface*. Oxford: Oxford University Press, 2007.

FÉRY, Caroline; ISHIHARA, Shinichiro (eds.). *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press, 2016.

FÉRY, Caroline; ISHIHARA, Shinichiro. Introduction. In: FÉRY, Caroline; ISHIHARA, Shinichiro (eds.). *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press, 2016, p. 1–18.

FIRBAS, Jan. On defining the theme in functional sentence analysis. *Linguistiques de Prague,* 1, 1964, 267-280.

FIRBAS, Jan. Functional sentence perspective in written and spoken communication. Cambridge: Cambridge University Press, 1992.

FIRENZUOLI, V. *Le forme intonative di valore illocutivo dell'Italiano parlato: analisi sperimentale di un corpus di parlato spontaneo (LABLITA)*. PhD dissertation. Università di Firenze, Firenze, 2003.

FIRENZUOLI, V.; SIGNORINI, S. L'unità informativa di topic: correlati intonativi. In: MAROTTA, G. (Ed.). *La coarticolazione*. Atti delle XIII giornate di studio del Gruppo di Fonetica Sperimentale, 28-30 nov. 2002. Pisa: ETS, 2003, p. 177–184.

FLEISS, Joseph L. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5), 1971.

FLORIDI, Luciano. *Information: A very short introduction*. Oxford: OUP, 2010.

FLORIDI, Luciano. Introduction. In: FLORIDI, LUCIANO (Org.). *The Routledge Handbook of Philosophy of Information*. London and New York: Routledge, 2016a, p. 1–4.

FLORIDI, Luciano. Semantic information. In: FLORIDI, LUCIANO (Org.). *The Routledge Handbook of Philosophy of Information*. London and New York: Routledge, 2016b, p. 44–49.

FLORIDI, Luciano. Semantic Conceptions of Information. *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), ZALTA, Edward N. (ed.), 2017. <https://plato.stanford.edu/archives/spr2017/entries/information-semantic/>.

GAMER, Matthias; LEMON, Jim; SINGH, Ian Fellows Puspendra. *irr: Various Coefficients of Interrater Reliability and Agreement*, version 0.84.1, R package, 2019. https://CRAN.R-project.org/package=irr

GIVÓN, Talmy. *Syntax: an introduction*. Vol. 1. Amsterdam/Philadelphia: John Benjamins Publishing, 2001a.

GIVÓN, Talmy. *Syntax: an introduction*. Vol. 2. Amsterdam/Philadelphia: John Benjamins Publishing, 2001b.

GOBBO, Oliver. *Marcadores discursivos em uma perspectiva informacional: análise prosódica e estatística*. M.A. Thesis. Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2019.

GUBIAN, M.; TORREIRA, F. J.; STRIK, H.; BOVES, L. W. J. Functional Data Analysis as a Tool for Analyzing Speech Dynamics A Case Study on the French Word c'était. *Fundamenta Informaticae* – FUIN, 2009, p. 2199–2202.

GUNDEL, Jeanette K. Universals of topic-comment structure. *Studies in syntactic typology*, 17(1), 1988, p. 209–239.

t' HART, J.; COLLIER, R.; COHEN, A. *A perceptual study on intonation: an experimental approach to speech melody*. Cambridge: Cambridge University Press, 1990.

HALLIDAY, M. A. K. Notes on transitivity and theme in English: Part 1. *Journal of Linguistics*, 3(1), 1967a, p. 37–81.

HALLIDAY, M. A. K. Notes on transitivity and theme in English: Part 2. *Journal of Linguistics*, 3(2), 1967b, p. 199–244.

HALLIDAY, M. A. K. Notes on transitivity and theme in English: Part 3. *Journal of Linguistics*, 4(2), 1968, p. 179–215.

HALLIDAY, M. A. K.; MATTHIESSEN, C. M. I. M. *Introduction to functional grammar*. Fourth edition. New York: Routledge, 2014 [1985].

HOWARD, David M.; ANGUS, Jamie. *Acoustics and psychoacoustics*. New York/London: Routledge, 2017.

HEIM, Irene. *The semantics of definite and indefinite noun phrases*. 1982. University of Massachusetts, Amherst, 1982.

HOCKETT, C. F. *A course in modern linguistics*. New York: MacMillan, 1958.

HOCKETT, C. F. The origin of speech. Scientific American, v. 203, p. 89–96, set. 1960.

JACKENDOFF, Ray S. *Semantic interpretation in generative grammar*. Cambridge: MIT Press, 1972.

KAMP, Hans; STOKHOF, Martin. Information in natural language. *Handbook of the Philosophy of Information*: 49-112. In: ADRIAANS, Pieter; BENTHEM, J. F. A. K. van. Philosophy of information. *Handbook of the Philosophy of Science* 8, 2008.

KARTTUNEN, Lauri. Discourse referents. In: MCCAWLEY, James D. (org.). Notes from the linguistic underground. New York: Academic Press, 1976.

KOENIG, Laura L.; LUCERO, Jorge C.; PERLMAN, Elizabeth. Speech production variability in fricatives of children and adults: Results of functional data analysis. *The Journal of the Acoustical Society of America* 124(5), 2008, p. 3158–60.

KRIFKA, Manfred. Basic notions of information structure. *Acta Linguistica Hungarica*, v. 55, n. 3-4, 2008 [2006], p. 243–276.

KRIFKA, Manfred; MUSAN, Renate. Information structure: overview and linguistic issues. In: KRIFKA, M.; MUSAN (eds.). *The Expression of Information Structure*. Berlin/Boston: De Gruyter Mouton, 2012, p. 1–49.

KUNO, Susumu. Functional sentence perspective: A case study from Japanese and English. *Linguistic inquiry* 3(3), 1972, p. 269–320.

KURODA, S-Y. The categorical and the thetic judgment: Evidence from Japanese syntax. *Foundations of language*, 1972, p. 153–185.

LAMBRECHT, Knud. On the status of SVO sentences in French discourse. Coherence and grounding in discourse, 1987, p. 217–261.

LAMBRECHT, K. *Information structure and sentence form: topic, focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press, 1994.

LANDIS, Richard j.; KOCH, Gary G. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 1977, p. 363–374.

LEE, SUNGBOK; BYRD, Dani; KRIVOKAPIĆ, Jelena. Functional data analysis of prosodic effects on articulatory timing. *The Journal of the Acoustical Society of America* 119(3), 2005, p. 1666-71.

LEHISTE, Ilse. Some factors affecting the duration of syllabic nuclei in English. *Proceedings of the first Salzburg conference on linguistics*, 1975, p. 81–104.

LI, C. N.; THOMPSON, S. A. Subject and topic: a new typology of language. In: LI, Charles N. (Ed.). *Subject and topic*. New York: Academic Press, 1976.

LIPKA, Leonhard. Functional sentence perspective, intonation, and the speaker. In: Gutknecht, Christoph (Ed.). *Grundbegriffe und Hauptströmungen der Linguistik*. Hoffmann und Campe, 1977, 133–141.

LUCERO, Jorge C.; KOENIG, Laura L. Time normalization of voice signals using functional data analysis. *The Journal of the Acoustical Society of America*, 108(4), 2000, p. 3158–60.

LÜDELING, Anke; RITZ, Julia ; STEDE, Manfred; AMIR, Zeldes. Corpus linguistics and information structure research. In: FÉRY, Caroline; ISHIHARA, Shinichiro (eds.). *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press, 2016, p. 599–620.

MACWHINNEY, Brian. *The CHILDES Project: tools for analyzing talk*. 3rd ed. Mahwah: Lawrence Erlbaum Associates, 2000.

MAIA ROCHA, B.; RASO, T. A unidade informacional de Introdutor Locutivo no Português Brasileiro: uma análise baseada em corpus. *Domínios de Linguagem*, 5(1), 2011, p. 327–343.

MAROGY, Amal E. *Kitāb Sībawayhi: Syntax and Pragmatics*. Leiden/Boston: Brill, 2010.

MARTIN, Philippe. WinPitch Corpus: A text-to-speech analysis and alignment tool. In: CRESTI, Emanuela; MASSIMO, Moneglia (eds.). *C-ORAL-ROM: integrated reference corpora for spoken Romance languages*. Amsterdam: John Benjamins, 2005, p. 40–51.

MARTÍNEZ, Carlota Nicolás; SOMACARRERA, Marina Lombán. MINI-CORPUS del español para IPIC. *Chimera: Romance Corpora and Linguistic Studies*, 5(2), 2018, p. 197–2015.

MATHESIUS, Vilém. *A functional analysis of Present Day English on a general Linguistic basis.* Paris/Prague: Mouton/Academia, 1975 [1961].

MATHESIUS, Vilém. Functional linguistics. In: VACHEK, Josef (ed.). *Pragma*. Amsterdam/Philadelphia: John Benjamins, 1983 [1929], p. 121-42.

MELLO; H. Methodological issues for spontaneous speech corpora compilation: the case of the C-ORAL-BRASIL. In: RASO, T.; MELLO, H (eds.). *Spoken corpora and linguistic studies*. Amsterdam: John Benjamins, 2014, p. 29–68.

MELLO; H., RASO, T.; MITTMANN, M.; VALE, H. P.; CÔRTES, P. O. Transcrição e segmentação prosódica do corpus C-ORAL-BRASIL: critérios de implementação e validação. In: RASO, T.; MELLO, H. (eds.). *C-ORAL-BRASIL I. Corpus de referência do português brasileiro falado informal*. Belo Horizonte: UFMG, 2012, p. 125–176.

MENDES SANTOS, S. *Parentético e outras unidades informacionais curtas: uma distinção através da análise das diferentes realizações prosódicas do lexema assim*. MA thesis. Facudade de Letras da Universidade Federal de Minas Gerais, 2020.

MEREU, Lunella (ed.). *Information structure and its interfaces*: Interface Explorations, vol. 19. Berlin: Walter de Gruyter, 2009.

MITHUN, Marianne. Modality and Mood in Iroquoian. In: NUYTS, J.; AUWERA, Johan van der (Eds.). *The Oxford Handbook of Modality and Mood*. Oxford: OUP, 2016.

MITTMANN, M. M. O C-ORAL-BRASIL e o estudo da fala informal: um novo olhar sobre o tópico no português brasileiro. 2012. 248 leaves. PhD dissertation, Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2012.

MITTMANN, Maryualê Malvessi; RASO, Tommaso. The C-Oral-Brasil Informationally Tagged Mini-Corpus. In: RASO, TOMMASO; PANUNZI, ALESSANDRO; MELLO, HELIANA R. (Orgs.). *Pragmatics and Prosody : Illocution, Modality, Attitude, Information Patterning and Speech Annotation*. Firenze: Firenze University Press, 2011. p. 151–183.

MONEGLIA, Massimo. Spoken Corpora and Pragmatics. *Revista Brasileira de Linguística Aplicada*, 11(2), 2011, p. 479–519.

MONEGLIA, Massimo; RASO, T.; MITTMANN, M. M.; MELLO, H. Challenging the perceptual relevance of prosodic breaks in multilingual spontaneous speech corpora: C-ORAL-BRASIL / C-ORAL-ROM in Speech Prosody 2010. In: *Satellite workshop on Prosodic Prominence: Perceptual, Automatic Identification.* Speech Prosody, 2010.

MONEGLIA, Massimo; RASO, Tommaso. Notes on Language into Act Theory (L-AcT). In: RASO, T.; MELLO, H. (Eds.). *Spoken corpora and linguistic studies*. Amsterdam: John Benjamins, 2014, p. 469–495.

MOLNÁR, Valéria; WINKLER, Susanne. Exploring the architecture of focus in grammar. In: MOLNÁR, Valéria; WINKLER, Susanne (eds.). *The Architecture of Focus*. Berlin: Mouton de Gruyter, 2005, p. 1–32.

MULLIGAN, K. *Mind, meaning and metaphysics: the philosophy and theory of language of Anton Marty*. Dordrecht/Boston/London: Kluwer Academic Publishers, 1990.

NESPOR, Marina; VOGEL, Irene. Prosodic phonology: with a new foreword. *Studies in Generative Grammar*. Vol. 28. Berlin/New York: Mouton de Gruyter, 2007.

NEWMEYER, Frederick J. *Language form and language function*. Cambridge/London: MIT press, 2000.

PANUNZI, A.; GREGORI, L. DB-IPIC: An XML database for the representation of information structure in spoken language. In: MELLO, H. R.; PANUNZI, A.; RASO, T. (Eds.). *Pragmatics and prosody: illocution, modality, attitude, information patterning and speech annotation*. Firenze: Firenze University Press, 2012, p. 133–150.

PANUNZI, A.; MITTMANN, M. M. The IPIC resource and a cross-linguistic analysis of information structure in Italian and Brazilian Portuguese. In: RASO, T.; MELLO, H. (Eds.). *Spoken corpora and linguistic studies*. Amsterdam: John Benjamins, 2014, p. 129–151.

PETROVA, S.; SOLF, M. On the methods of information-structural analysis in historical texts. A case study on Old High German. *New Approaches to Word Order Variation and Word Order Change*. Berlin and New York: de Gruyter, 2009, p. 121–160.

PING, Alvin Leong. *Theme and rheme:* an alternative account. European university studies, vol. 266. Bern: Peter Lang, 2004.

POLLAR, C.; SAG, I. A. *Information-Based Syntax and Semantics, Vol. 1: Fundamentals*. Chicago: University of Chicago Press, 1987.

PRINCE, Ellen. The ZPG letter: Subjects, definiteness, and information-status. In: MANN, William C.; THOMPSON, Sandra A. (Eds.). *Discourse Description: Diverse linguistic analyses of a fund-raising text*. Philadelphia: John Benjamins, 1992, p. 295–326.

R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. Available at: http://www.R-project.org. Accessed: January 2, 2020.

RAMOS, A. C. *The use of certainty adverbs in Brazilian Portuguese and American English: a semantic/pragmatic approach*. 2015. 128 ff. MA thesis. Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2016.

RAMSAY, James O.; SILVERMAN, B. W. *Functional Data* Analysis. New York: Springer-Verlag, 2005.

RAMSAY, James O.; DALZELL, C. J. Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1991, p. 539–572.

RASO, Tommaso. Fala e escrita: meio, canal, consequências pragmáticas e linguísticas. *Domínios da Linguagem*, 7(2), 2013 p. 12–46.

RASO, Tommaso. Prosodic constraints for discourse markers. In: RASO, T.; MELLO, H. (Eds.). *Spoken corpora and linguistic studies*. Amsterdam: John Benjamins, 2014, p. 411–467.

RASO, Tommaso; BOSSAGLIA, G. The C-ORAL IT minicorpus: informational annotation and acoustic quality. In preparation.

RASO, Tommaso; CAVALCANTE, F.; MITTMANN, M. Prosodic forms of the Topic information unit in a cross-linguistic perspective: a first survey. *Proceedings of the SLI-GSCP International Conference*, 13-15 June, 2016. Anna de Meo & Francesca Maria Dovetto. Rome: Aracne editrice, 2017.

RASO, Tommaso; MELLO, Heliana (eds.). *C-ORAL-BRASIL I: corpus de referência de português brasileiro falado informal*. Belo Horizonte: UFMG, 2012.

RASO, Tommaso; MELLO, H. R.; FERRARI, L. A. The C-ORAL-BRASIL II: corpora of Brazilian Portuguese speech in formal, media, and telephonic interactions, forthcoming.

RASO, Tommaso; MITTMANN, Maryualê. Validação estatística dos critérios de segmentação da fala espontânea no corpus C-ORAL-BRASIL. *Revista de Estudos da Linguagem*, 17(2), 2009, p. 73–91.

RASO, Tommaso; MITTMANN, M. As principais medidas da fala. In.: RASO, T; MELLO, H. (org.). *C-ORAL-BRASIL I: Corpus de referência do Português Brasileiro falado informal*. Belo Horizonte: Editora UFMG, 2012, p. 177–222.

RASO, Tommaso; MITTMANN, M. M.; MENDES, A. C. O. O papel da pausa na segmentação prosódica de corpora de fala. *Revista de Estudos da Linguagem*, 23(3), 2015.

RASO, T.; VIEIRA, M. A description of Dialogic Units/Discourse Markers in spontaneous speech corpora based on phonetic parameters. *Chimera: Romance Corpora and Linguistic Studies*, 3, 2016, p. 221–249. Available at: <https://revistas.uam.es/index.php/chimera/issue/archive>.

REINHART, Tanya. Pragmatics and linguistics: an analysis of sentence topics. *Philosophica*, 27(1), 1981, p. 53–94.

REINHART, Tanya. *Pragmatics and linguistics: an analysis of sentence topics*. Bloomington: Indiana Linguistics Club, 1982.

ROCHA, B. N. *Características prosódicas do tópico em PE e o uso do pronome lembrete*. 2012. 264 ff. MA thesis, Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2012.

ROCHA, B. N. *A forma prosódica da ilocução de Ordem em Português Brasileiro e Italiano: proposta de metodologia baseada em corpus*. PhD dissertation, Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2016.

ROCHA, Bruno Neves Rati de Melo. O corpus C-ORAL-ESQ e a estrutura informacional da fala de pacientes com esquizofrenia. *Working papers em Linguística*, 20, p. 212–238, 2019.

ROOTH, Mats. Alternative semantics. In: FÉRY, Caroline; ISHIHARA, Shinichiro (eds.). *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press, 2016, p.19–40.

SASSE, Hans-Jürgen. The thetic/categorical distinction revisited. *Linguistics*, 25(3), 1987, p. 511–580.

SCARANO, A. A The prosodic annotation of C-ORAL-ROM and the structure of information in spoken language. In: MEREU, L (ed.). *Information structures and its interfaces*. Berlin: Mouton de Gruyter; 2009, p. 51–74.

SEARLE, John R. *A taxonomy of illocutionary acts*. London: Cambridge University Press, 1975.

SEARLE, John R. *An essay in the philosophy of language*. London: Cambridge University Press, 1979.

SHANNON, Claude Elwood. A mathematical theory of communication. *Bell system technical journal*, 27(3), 1948, p. 379–423.

SHANNON, Claude Elwood; WEAVER, W. *The Mathematical Theory of Communication*. Urbana: University of Illinois Press, 1949.

SHANNON, Claude Elwood. A Mathematical Theory of Communication. In: WYNER, A. D.; SLOANE, Neil (eds.). *Claude E. Shannon: Collected Papers*. New York: Wiley & Sons, 1993 [1950], p. 5–83.

SIGNORINI, S. *Topic e soggetto in corpora di italiano parlato spontaneo*. Phd. Dissertation, Università di Firenze, 2005.

SILVA, Andressa C. Oliveira; FONSECA, Aline Alves. Prosody and processing: comprehension and production of topic-comment and subject-predicate structures in Brazilian Portuguese. *Revista Estudos da Linguagem*, 26(4), Belo Horizonte, 2018, p. 1601–46.

SILVA, Luis Filipi. *Sintaxe da fala, probabilidade e cognição: Uma proposta integrada de investigação em estudo sobre NPs baseado em corpus*. PhD dissertation. Faculdade de Letras, Universidade Federal de Minas Gerais, 2020.

STALNAKER, Robert. Pragmatics. *Synthese,* 22(1/2), 1970, p. 272–289.

STALNAKER, Robert. Pragmatic presuppositions. In: MUNITZ, M. K.; UNGER, P. (Eds.). *Semantics and Philosophy*. New York: New York University Press, 1974, p. 197-213.

STALNAKER, Robert. Assertion. In: COLE, PETER (org.). *Pragmatics*, 9. New York: Academic Press, 1978, p. 315–332.

STALNAKER, Robert. *Context*. Oxford: Oxford University Press, 2014.

STRAWSON, P. F. Identifying reference and truth-values. *Theoria*, 30(2), Blackwell Publishing Ltd, 1964, p. 96–118.

TEIXEIRA FALCÃO, Bárbara Helohá Falcão. *Correlatos fonético-acústicos de fronteiras prosódicas na fala espontânea*. 2018. 174 leaves. MA thesis. Belo Horizonte: Faculdade de Letras, Universidade Federal de Minas Gerais, 2018.

TEIXEIRA, Bárbara Helohá Falcão; MALVESSI MITTMANN, Maryualê. Acoustic Models for the Automatic Identification of Prosodic Boundaries in Spontaneous Speech. *Revista de Estudos da Linguagem*, 26(4), 2018, p. 1455–1488.

t'HART, J. T.; COLLIER, R.; COHEN, A. *A perceptual study on intonation: an experimental approach to speech melody*. Cambridge: CUP, 1990.

TUCCI, Ida. L'inciso: caratteristiche morfosintattiche e intonative in un corpus di riferimento. In: LEONI, F. A. *et al.* (org.). Il parlato italiano. *Atti del Convegno Nazionale*. Napoli: M. D'Auria, 2004, p. 1–14.

TUCCI, Ida. "Obiter dictum". La funzione informativa delle unità parentetiche. In: PETTORINO, M.; GIANNINI, A.; DOVETTO, F. (org.). La comunicazione parlata, 3, 2009, Napoli. *Atti del Convegno Internazionale GSCP*, 1. Napoli: Università degli Studi di Napoli l'Orientale, 2010, p. 635–654.

VACHEK, Josef. On the integration of the peripheral elements into the system of language. *Travaux linguistiques de Prague*, 2, 1966, p. 23–37.

VALLDUVÍ, Enric. *The Informational Component*. New York: Garland, 1992.

VALLDUVÍ, Enric; ENGDAHL, Elisabet. The linguistic realization of information packaging. *Linguistics*, 34(3), 1996, p. 459–520.

VIERA, Anthony J.; GARRETT, Joanne M. Understanding interobserver agreement: the kappa statistic. *Fam med*, 37 (5), 2005, p. 360–363.

WASSERSTEIN, Ronald L.; LAZAR, Nicole A. The ASA statement on p-values: context, process, and purpose. *The American Statistician*, 20(2), 2016, p. 129–133.

WEAVER, Warren. The mathematics of communication. *Scientific American*, 181(1), 1949, p. 11–15.

WEIL, Henri. *The order of words in the ancient languages compared with that of the modern languages*. Amsterdam: John Benjamins, 1978.

Appendix

APPENDIX A – Agreement test protocol

Source: author

TESTE DE ACORDO NA IDENTIFICAÇÃO DE TOP
PROTOCOLO PARA ANOTAÇÃO

LEIA TODO ESTE PROTOCOLO ANTES DE DAR INÍCIO À TAREFA

1) Programa a ser utilizado para a tarefa
   - Praat.

   IMPORTANTE: antes de dar início à tarefa, desativar a visualização da curva de *pitch* (com o áudio aberto, clicar em "Pitch" e então em "Show Pitch". Isso fará com que a curva azul sobreposta ao espectrograma não seja mais exibida).

2) Arquivos

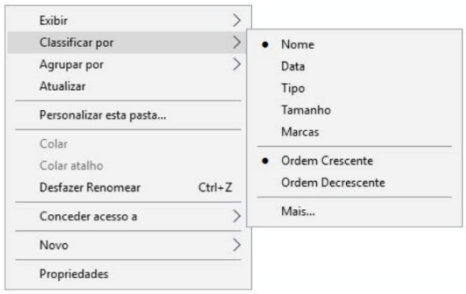
   (a) 500 arquivos wav:
   - 100 enunciados-alvo: 40 da parte de *media*, 40 de *natural context*, 20de*telephonic*.
   - 400 enunciados-contexto (dois antes e dois depois de cada enunciado alvo)

   (b) Um arquivo de Excel com três planilhas, uma para cada seção (*media*, *natural contex*, *telephonic*)

   (c) Exemplo de como os dados estão nas planilhas:

bnatbu03
bnatbu03_utterance_229.wav
bnatbu03_utterance_230.wav
bnatbu03_utterance_231ALVO.wav
bnatbu03_utterance_232.wav
bnatbu03_utterance_233.wav

   Observações:
   - Os arquivos de áudio com os enunciados a serem etiquetados estão sempre com "ALVO" como parte do nome
   - os arquivos de áudio serão exibidos em ordem, contanto que a organização da pasta esteja por classificada por Nome/Ordem Crescente, como a imagem abaixo mostra.



3) As etiquetas

(a) =COM=, =COB=, ou =CMM=: para unidades ilocucionárias, conforme a natureza da unidade
(b) =TOP=: para unidades de tópico; etiqueta usada tanto para tópicos comuns, quanto para tópicos em lista
(c) =UNC=:para usar quando, com grau alto de certeza, acharmos que a quebra anotada não existe. Caso isso acontece mais de 3 ou 4 vezes, comunicar aos demais.
(d) =NTP=: para demais unidades, não importando se são textuais, dialógicas ou mesmo vazias de valor informacional (i.e., SCA, TMT, e EMP)

4) O processo de etiquetagem

A ser feita na planilha fornecida apenas nos enunciados sinalizados em verde e cujos arquivos wav contenham “ALVO” no nome.

(a) Etapa 1: anotar as unidades ilocucionárias no enunciado-alvo (=COM=, =COB=, ou =CMM=)
(b) Etapa 2: anotar as unidades de TOP no enunciado-alvo (=TOP=), lembrando que o tópico configura-se como o âmbito cognitivo da aplicação de uma unidade ilocucionárias e que a etiqueta UNC poderá ser usada nos casos especificados em (3.c) acima
(c) Etapa 3: etiquetar as unidades prosódicas restantes no enunciado-alvo (=NTP=)

5) Notas para guiar a identificação dos tópicos:

(a) Definição funcional: o tópico é a unidade informacional que estabelece *um domínio cognitivo de identificação* em relação ao qual a ilocução deve ser interpretada.
(b) As formas prosódicas devem ser consideradas como sinais formais da realização da unidade, sendo, pois, necessárias mas não suficientes para determinar a identidade informacional de uma unidade candidata a tópico.
(c) Princípio norteador da etiquetagem: não multiplicar TOPs, ou seja, não etiquetar como TOP casos potencialmente problemáticos e ambíguos, tendo em mente que:

i. O TOP é sempre referencial, devendo, portanto, sempre se referir a um domínio de identificação, seja este domínio individual, de classe ou conceitual

ii. Unidades preenchidas lexicalmente só por pronomes devem ser etiquetadas como TOP somente em casos em que não houver dúvida. O critério para solucionar a dúvida é principalmente o prosódico, sendo preciso que a proeminência seja claramente percebida
iii. A prosódia é importante, mas o critério semântico, que estabelece a natureza da unidade, deve ser sempre considerado; para ser TOP, a unidade deve passar tanto pelo critério semântico quanto pelo prosódico
iv. Itens como "aí" e "então" só poderão preencher sozinhos um TOP se forem referenciais (e.g., se apresentarem significado temporal claro ou então se forem anafóricos). Nesses casos, o critério semântico é mais forte. Se houver dúvida sobre este critério, não etiquetar como tópico.
v. Para atribuir valor anafórico a expressões como "então" e "aí", em que a anáfora não é explícita e em que é frequente e sempre possível uma interpretação não anafórica, os motivos devem ser muito fortes. Portanto, não há dúvida que um TOP como "DADO ISSO /=TOP= FAREMOS ASSIM //=COM=" é um TOP anafórico; mas deve-se tomar muito cuidado em casos como "ENTÃO /=TOP= FAREMOS ASSIM //=COM=". Para que atribuirmos valor anafórico a expressões que não sejam exclusiva e explicitamente anafóricas, precisamos de um ótimo motivo e de uma forma prosódica muito forte. Na dúvida, consideramos essas formas dialógicas, etiquetando-as como =NTP=.
vi. Se o conteúdo locutivo for muito reduzido, a forma prosódica tem que ser muito clara para que a unidade seja etiquetada como TOP

Quando um TOP é seguido por uma unidade exibindo um claro contraste de nível prosódico, considerar com cuidado especial se a unidade não deve ser considerada um PAR (ou seja, etiquetada como NTP), mesmo se na sua porção final houver um movimento que lembre um núcleo de tópico.

APPENDIX B – R script for sampling utterances
to be used in the agreement test

Source: author

```
#This script randomly samples utterances from the C-ORAL-BRASIL corpus to be used in an interrater agreement test (Fleiss' Kappa) for TOP detection.
#Output: 3 csv files for each section of the corpus (X = section): (1) dataX.csv: the utterances and their contexts (40 from media, 40 from natural context, 20 from telephonic); (2) indexX.csv: file name and utterances rank; (3) sampleX.csv: name of files from which samples are drawn. File (3) is redundant, as its information is available in (2).

#Part 1
#This script has individual sections for each of the three sections of the C-ORAL-BR II, i.e. natural context, media, and telephonic. Only one of such sections is shown here
#An individual part can be run independently from the others, but PART 1 has to be run before any of them.

#Txt files have to be separated by corpus section in folders named "med", "nat,", and "tel", and these folders must be together in the same folder, which can be named as desired.
# The path to the folder is the only information the user has to provide

path.to.folders <- "/Users/_Kappa TOP/kappa top sampling/4_complexos SEM interromp"

setwd(path.to.folders)
folders<-dir(path.to.folders)
#creates columnNames
columnNames.a<-c("file", "\n")
columnNames.b<- c("file", "context 1","context 2","Utt","context 3","context 4", "\n")
columnNames.c<- c("file", "utt-2","utt-1","utt","utt+1","utt+2", "\n")

#end of Part 1

#Natural context
#gets NAT CONTEXT texts to be sampled (creates a file with name of files selected)
#Preparation
setwd(path.to.folders)
file.nat<- dir(paste(path.to.folders,"/nat", sep=""))
sample.nat<- paste(path.to.folders, "/sampleNat.csv", sep="")
data.nat<- paste(path.to.folders, "/dataNat.csv", sep="")
index.nat<- paste(path.to.folders, "/indexNat.csv", sep="")
cat(columnNames.a, file=sample.nat, sep=",", append=TRUE)
cat(columnNames.b, file=data.nat, sep=",", append=TRUE)
cat(columnNames.b, file=index.nat, sep=",", append=TRUE)

#gets texts with more than 5 utterances

setwd(paste(path.to.folders,"/nat",sep=""))
for (i in 1:length(file.nat)) {
```

```
		text <-scan(file.nat[i], what="char", sep="$", quiet=TRUE, encoding="UTF-
		8", quote="")
		text.1<- text[text!=""] #gets rid of "empty" elements \
		text.2<- text.1[text.1!=" "] #gets rid of " " elements \
		size.text<- length(text.2)
		if (size.text>5){
			cat(file.nat[i], "\n", file=sample.nat, append=T)
		}}

#gets 40 random texts
selected.texts.1<-scan(sample.nat, what="char", sep="\n", quiet=TRUE, encoding="UTF-8",
quote="")
selected.texts.1a<- length(selected.texts.1)
selected.texts.2<-sort(sample(1:selected.texts.1a,40))

#gets the utterances and their contexts (2 preceding and 2 following utts)
for (i in selected.texts.2) {
	text <-scan(file.nat[i], what="char", sep="$", quiet=TRUE, encoding="UTF-8",
quote="")
	text.1<- text[text!=""] #gets rid of "empty" elements
	text.2<- text.1[text.1!=" "] #gets rid of " " elements
		size.text<- length(text.2)
		sample.range<- 3:(size.text-2)
		selected.utt.index<- sample(sample.range,1)
		selected.utt<-text.2[selected.utt.index]
		#checks if selected.utt is concluded and complex
		complex.concluded<- grepl("//", selected.utt, fixed=T) & grepl(" / ",
		selected.utt, fixed =T)

		while (complex.concluded==FALSE) {
		selected.utt.index<- sample(sample.range,1)
		selected.utt<-text.2[selected.utt.index]
		complex.concluded<- grepl("//", selected.utt, fixed=T) & grepl(" / ",
selected.utt, fixed =T)
		}
		context.1<- text.2[selected.utt.index-2]
		context.2<- text.2[selected.utt.index-1]
		context.3<- text.2[selected.utt.index+1]
		context.4<- text.2[selected.utt.index+2]
		cat(file.nat[i], context.1, context.2, selected.utt, context.3, context.4, "\n",
file=data.nat, sep=",", append=TRUE)
		cat(file.nat[i],selected.utt.index-2,selected.utt.index-
1,selected.utt.index,selected.utt.index+1, selected.utt.index+2 , "\n", file=index.nat, sep=",",
append=TRUE)
}
#Ends Nat context
```

APPENDIX C – Instruction for prosodic segmentation revision(agreement-test utterances)

Source: author

REVISÃO DA SEGMENTAÇÃO PROSÓDICA
ENUNCIADOS AMOSTRADOS PARA O TESTE DE ACORDO NA IDENTIFICAÇÃO DE TOP

Descrição dos dados e da tarefa

1. Materiais

Os arquivos para revisão estão organizados em 3 pastas, uma para cada parte do corpus (media, natural context e telephonic).

Pasta *Utterances_MED*:

- 200 arquivos .wav de media (40 enunciado-alvo, 160 enunciados-contexto)
- 1 arquivo .csv contendo as transcrições dos enunciados-alvo e -contexto

Pasta *Utterances_NAT*:

- 200 arquivos .wav de nat. cont. (40 enunciado-alvo, 160 enunciados-contexto)
- 1 arquivo .csv contendo as transcrições dos enunciados-alvo e -contexto

Pasta *Utterances_TEL*:

- 100 arquivos .wav de tel (20 enunciado-alvo, 80 enunciados-contexto)
- 1 arquivo .csv contendo as transcrições dos enunciados-alvo e -contexto

2. Organização dos áudios

Para cada enunciado sorteado (“enunciado-alvo”), foram salvos outros quatro enunciados (“enunciados-contexto”), dois anteriores e dois subsequentes ao alvo. Assim, para cada linha no arquivo .csv há 5 arquivos .wav. O enunciado-alvo será sempre o terceiro arquivo, como sinalizado na lista abaixo:

bnatbu03_utterance_229.wav *enunciado-contexto*
bnatbu03_utterance_230.wav *enunciado-contexto*
bnatbu03_utterance_231.wav *enunciado-alvo*
bnatbu03_utterance_232.wav *enunciado-contexto*
bnatbu03_utterance_233.wav *enunciado-contexto*

3. Tarefa

Revisar a segmentação prosódica dos enunciados-alvo

APPENDIX D – Praat script for getting pitch contours

Source: author

```
# Gets time-stamped pitch contours from wav files in a folder and store them in separate text
files
### Wav files have to be properly named:
##### 1st characters must be "M" for male speaker; any other character is interpreted as
female
##### 2-4 initial chars. identify the hypothesized top type (1st, 2nd, 3rd)

clearinfo
form Time-stampled f0; in most cases values should be kept as is.
 positive StartType 2
 positive StartTypeb 1
 word FileOut _List.txt
 word FileOut2 _List_sex.txt
 word AudiofileExtension *.wav
endform

Create Strings as file list... list 'audiofileExtension$'
numberOfFiles = Get number of strings


if !numberOfFiles
        exit There are no sound files in the folder!
endif

filedelete 'fileOut$'
fileappend 'fileOut$' filename type 'newline$'

filedelete 'fileOut2$'
fileappend 'fileOut2$' sex&Toptype 'newline$'

for ifile from 1 to numberOfFiles

 select Strings list
 audiofile$ = Get string: 'ifile'

 Read from file... 'audiofile$'
 type$ = mid$(audiofile$,'startType',3)
 sex$ = mid$(audiofile$,'startTypeb',2)

 fileappend 'fileOut2$' 'sex$' 'newline$'

if sex$ = "M*"

filename$ = selected$("Sound")
 To Pitch: 0, 75, 300
```

```
 Down to PitchTier
 Save as headerless spreadsheet file: 'audiofile$' - "wav" + "txt"

else
filename$ = selected$("Sound")
 To Pitch: 0, 100, 500
 Down to PitchTier
 Save as headerless spreadsheet file: audiofile$ - "wav"+"txt"

        fileappend 'fileOut$' 'filename$"tab$"type$' 'newline$'

endif
endfor
select all
Remove
Print Task completed
```

APPENDIX E – Praat script for creating landmark tables

Source: author

```
#creates landmark tables
#curves with two relevant f0 movements are saved separate from those with one relevant f0 movement
#vowels are must be labled "v1" and "v2" when thare are two relevant movements and "v" when there is only one
clearinfo
form Landmaks
 word FileOut LandmarksOneVowel.txt
 word FileOut1 LandmarksTwoVowels.txt
 word TextfileExtension *.TextGrid
endform

Create Strings as file list... textlist 'textfileExtension$'
numberOfFiles = Get number of strings

if !numberOfFiles
  exit Required files not in the folder!
endif
filedelete 'fileOut$'
fileappend 'fileOut$' filename type sex v1_beg v1_end'newline$'
filedelete 'fileOut1$'
fileappend 'fileOut1$' filename type sex v1_beg v1_end v2_beg v2_end'newline$'

beg_v=0
end_v=0
beg_v1=0
end_v1=0
beg_v2=0
end_v2=0

for ifile from 1 to numberOfFiles

select Strings textlist
textfile$ = Get string: 'ifile'

  Read from file... 'textfile$'

  numOfintervals= Get number of intervals: 1
  type$ = mid$(textfile$,2,3)
  sex$ = mid$(textfile$,1,1)
  filename$=textfile$-".TextGrid"

  for interval from 1 to numOfintervals
                label$= Get label of interval: 1, interval
```

```
                    if label$ == "v"
                        beg_v = Get starting point: 1, interval
                        end_v = Get end point: 1, interval
        fileappend 'fileOut$' 'filename$' 'type$' 'sex$' 'beg_v' 'end_v' 'newline$'
                        endif

                        if label$ == "v1"
                        beg_v1 = Get starting point: 1, interval
                        end_v1 = Get end point: 1, interval
                        endif

                        if label$ == "v2"
                        beg_v2 = Get starting point: 1, interval
                        end_v2 = Get end point: 1, interval
                        fileappend 'fileOut1$' 'filename$' 'type$' 'sex$' 'beg_v1' 'end_v1'
'beg_v2' 'end_v2' 'newline$'
                        endif
    endfor
    endfor

    select all
    Remove
    print Task completed!
```