

REVISTA DE ESTUDOS DA LINGUAGEM

Faculdade de Letras da UFMG

ISSN

Impresso: 0104-0588

On-line: 2237-2083

V.26 - N° 4



REVISTA DE ESTUDOS DA LINGUAGEM

Universidade Federal de Minas Gerais

REITORA: Sandra Regina Goulart Almeida
VICE-REITOR: Alessandro Fernandes Moreira

Faculdade de Letras:

DIRETORA: Graciela Inés Ravetti de Gómez
VICE-DIRETORA: Sueli Maria Coelho

Editora-chefe

Heliana Ribeiro de Mello

Editores-associados

Aderlande Pereira Ferraz (UFMG)
Gustavo Ximenes Cunha (UFMG)
Maria Cândida Trindade Costa de Seabra (UFMG)

Revisão e Normalização

Alda Lopes Durães Ribeiro
Heliana Ribeiro de Mello

Editoração eletrônica

Alda Lopes Durães Ribeiro
Henrique Vieira da Silva
Úrsula Francine Massula

Secretaria

Úrsula Francine Massula

REVISTA DE ESTUDOS DA LINGUAGEM, v.1 - 1992 - Belo Horizonte, MG,
Faculdade de Letras da UFMG

Histórico:

1992 ano 1, n.1 (jul/dez)
1993 ano 2, n.2 (jan/jun)
1994 Publicação interrompida
1995 ano 4, n.3 (jan/jun); ano 4, n.3, v.2 (jul/dez)
1996 ano 5, n.4, v.1 (jan/jun); ano 5, n.4, v.2; ano 5, n. esp.
1997 ano 6, n.5, v.1 (jan/jun)

Nova Numeração:

1997 v.6, n.2 (jul/dez)
1998 v.7, n.1 (jan/jun)
1998 v.7, n.2 (jul/dez)

1. Linguagem - Periódicos I. Faculdade de Letras da UFMG, Ed.

CDD: 401.05

ISSN: Impresso: 0104-0588
 On-line: 2237-2083

REVISTA DE ESTUDOS DA LINGUAGEM

V. 26 - Nº 4 - out.-dez. 2018

Indexadores

Diadorim [Brazil]

DOAJ (Directory of Open Access Journals) [Sweden]

DRJI (Directory of Research Journals Indexing) [India]

EBSCO [USA]

JournalSeek [USA]

Latindex [Mexico]

Linguistics & Language Behavior Abstracts [USA]

MIAR (Matriu d'Informació per a l'Anàlisi de Revistes) [Spain]

MLA Bibliography [USA]

OAJI (Open Academic Journals Index) [Russian Federation]

Portal CAPES [Brazil]

REDIB (Red Iberoamericana de Innovación y Conocimiento Científico) [Spain]

Sindex (Scientific Indexing Services) [USA]

Web of Science [USA]

WorldCat / OCLC (Online Computer Library Center) [USA]

ZDB (Elektronische Zeitschriftenbibliothek) [Germany]



REVISTA DE ESTUDOS DA LINGUAGEM

Editora-chefe

Heliana Ribeiro de Mello (UFMG, Belo Horizonte/MG, Brasil)

Organizadores do número

Plínio A. Barbosa (Unicamp, Campinas/SP, Brasil)

Tommaso Raso (UFMG, Belo Horizonte/MG, Brasil)

Editores-associados

Aderlande Pereira Ferraz (UFMG, Belo Horizonte/MG, Brasil)

Gustavo Ximenes Cunha (UFMG, Belo Horizonte/MG, Brasil)

Maria Cândida Trindade Costa de Seabra (UFMG, Belo Horizonte/MG, Brasil)

Conselho Editorial

Alejandra Vitale (UBA, Ciudad Autónoma de Buenos Aires, Argentina)

Didier Demolin (Université de la Sorbonne Nouvelle Paris 3, Paris, França)

Ieda Maria Alves (USP, São Paulo/SP, Brasil)

Jairo Nunes (USP, São Paulo/SP, Brasil)

Scott Schwenter (OSU, Columbus, Ohio, Estados Unidos)

Shlomo Izre'el (TAU, Tel Aviv, Israel)

Stefan Gries (UCSB, Santa Barbara/CA, Estados Unidos)

Teresa Lino (NOVA, Lisboa, Portugal)

Tjerk Hagemeijer (ULisboa, Lisboa, Portugal)

Comissão Científica

Aderlande Pereira Ferraz (UFMG, Belo Horizonte/MG, Brasil)
Alessandro Panunzi (UniFl, Florença, Itália)
Alina M. S. M. Villalva (ULisboa, Lisboa, Portugal)
Aline Alves Ferreira (UCSB, Santa Barbara/CA, Estados Unidos)
Ana Lúcia de Paula Müller (USP, São Paulo/SP, Brasil)
Ana Maria Carvalho (UA, Tucson/AZ, Estados Unidos)
Anabela Rato (U of T, Toronto/ON, Canadá)
Aparecida de Araújo Oliveira (UFV, Viçosa/MG, Brasil)
Aquiles Tescari Neto (UNICAMP, Campinas/SP, Brasil)
Augusto Soares da Silva (UCP, Braga, Portugal)
Beth Brait (PUC-SP/ Universidade de São Paulo-USP, São Paulo/SP, Brasil)
Bruno Neves Rati de Melo Rocha (UFPA, Altamira/PA, Brasil)
Carmen Lucia Barreto Matzenauer (UCPEL, Pelotas/RS, Brasil)
César Nardelli Cambraia (UFMG, Belo Horizonte/MG, Brasil)
Cristina Name (UFJF, Juiz de Fora/MG, Brasil)
Charlotte C. Galves (UNICAMP, Campinas/SP, Brasil)
Deise Prina Dutra (UFMG, Belo Horizonte/MG, Brasil)
Diana Luz Pessoa de Barros (USP/ UPM, São Paulo/SP, Brasil)
Dylia Lysardo-Dias (UFSJ, São João del-Rei/MG, Brasil)
Edwiges Morato (UNICAMP, Campinas/SP, Brasil)
Emília Mendes Lopes (UFMG, Belo Horizonte/MG, Brasil)
Esmeralda V. Negrão (USP, São Paulo/SP, Brasil)
Flávia Azeredo Cerqueira (JHU, Baltimore/MD, Estados Unidos)
Gabriel de Avila Othero (UFRGS, Porto Alegre/RS, Brasil)
Gerardo Augusto Lorenzino (TU, Filadélfia/PA, Estados Unidos)
Gláucia Muniz Proença de Lara (UFMG, Belo Horizonte/MG, Brasil)
Hanna Batoréo (UAb, Lisboa, Portugal)
Heliana Ribeiro de Mello (UFMG, Belo Horizonte/MG, Brasil)
Heronides Moura (UFSC, Florianópolis/SC, Brasil)
Hilario Bohn (UCPEL, Pelotas/RS, Brasil)
Hugo Mari (PUC-Minas, Belo Horizonte/MG, Brasil)
Ida Lucia Machado (UFMG, Belo Horizonte/MG, Brasil)
Ieda Maria Alves (USP, São Paulo/SP, Brasil)
Ivã Carlos Lopes (USP, São Paulo/SP, Brasil)
Jairo Nunes (USP, São Paulo/SP, Brasil)
Jean Cristtus Portela (UNESP-Araraquara, Araraquara/SP, Brasil)

João Antônio de Moraes (UFRJ, Rio de Janeiro/ RJ, Brasil)
João Miguel Marques da Costa (Universidade Nova da Lisboa, Lisboa, Portugal)
João Queiroz (UFJF, Juiz de Fora/MG, Brasil)
José Magalhaes (UFU, Uberlândia/MG, Brasil)
João Saramago (Universidade de Lisboa)
José Borges Neto (UFPR, Curitiba/PR, Brasil)
Kanavillil Rajagopalan (UNICAMP, Campinas/SP, Brasil)
Laura Alvarez Lopez (Universidade de Estocolmo, Stockholm, Suécia)
Leo Wetzels (Free Univ. of Amsterdam, Amsterdã, Holanda)
Laurent Filliettaz (Université de Genève, Genebra, Suíça)
Leonel Figueiredo de Alencar (UFC, Fortaleza/CE, Brasil)
Livia Oushiro (UNICAMP, Campinas/SP, Brasil)
Lodenir Becker Karnopp (UFRGS, Porto Alegre/RS, Brasil)
Lorenzo Teixeira Vitral (UFMG, Belo Horizonte/MG, Brasil)
Luiz Amaral (UMass Amherst, Amherst/MA, Estados Unidos)
Luiz Carlos Cagliari (UNESP, São Paulo/SP, Brasil)
Luiz Carlos Travaglia (UFU, Uberlândia/MG, Brasil)
Marcelo Barra Ferreira (USP, São Paulo/SP, Brasil)
Marcia Cançado (UFMG, Belo Horizonte/MG, Brasil)
Márcio Leitão (Universidade Federal da Paraíba, João Pessoa/PB, Brasil)
Marcus Maia (UFRJ, Rio de Janeiro/RJ, Brasil)
Maria Antonieta Amarante M. Cohen (UFMG, Belo Horizonte/ MG, Brasil)
Maria Bernadete Marques Abaurre (UNICAMP, Campinas/SP, Brasil)
Maria Cecília Camargo Magalhães (PUC-SP, São Paulo/SP, Brasil)
Maria Cecília Magalhães Mollica (UFRJ, Rio de Janeiro/RJ, Brasil)
Maria Cândida Trindade Costa de Seabra (UFMG, Belo Horizonte/MG, Brasil)
Maria Cristina Figueiredo Silva (UFPR, Curitiba/PR, Brasil)
Maria do Carmo Viegas (UFMG, Belo Horizonte/MG, Brasil)
Maria Luíza Braga (PUC/RJ, Rio de Janeiro/RJ, Brasil)
Maria Marta P. Scherre (UNB, Brasília/DF, Brasil)
Miguel Oliveira, Jr. (Universidade Federal de Alagoas)
Milton do Nascimento (PUC-Minas, Belo Horizonte/MG, Brasil)
Monica Santos de Souza Melo (UFV, Viçosa/MG, Brasil)
Patricia Matos Amaral (UI, Bloomington/IN, Estados Unidos)
Paulo Roberto Gonçalves Segundo (USP, São Paulo/SP, Brasil)
Philippe Martin (Université Paris 7, Paris, França)
Rafael Nonato (Museu Nacional-UFRJ, Rio de Janeiro/RJ, Brasil)
Raquel Meister Ko. (Freitag, UFS, Brasil)

Roberto de Almeida (Concordia University, Montreal/QC, Canadá)
Ronice Müller de Quadros (UFSC, Florianópolis/SC, Brasil)
Ronald Beline (USP, São Paulo/ SP, Brasil)
Rove Chishman (UNISINOS, São Leopoldo/RS, Brasil)
Sanderléia Longhin-Thomazi (UNESP, São Paulo/SP, Brasil)
Sergio de Moura Menuzzi (UFRGS, Porto Alegre/RS, Brasil)
Seung- Hwa Lee (UFMG, Belo Horizonte/MG, Brasil)
Sírrio Possenti (UNICAMP, Campinas/SP, Brasil)
Suzi Lima (U of T / UFRJ, Toronto/ON - Rio de Janeiro/RJ, Brasil)
Thais Cristofaro Alves da Silva (UFMG, Belo Horizonte/MG, Brasil)
Tommaso Raso (UFMG, Belo Horizonte/MG-Brasil)
Tony Berber Sardinha (PUC-SP, São Paulo/SP, Brasil)
Ubiratã Kickhöfel Alves (UFRGS, Porto Alegre/RS, Brasil)
Vander Viana (University of Stirling, Stirling/Sld, Reino Unido)
Vanise Gomes de Medeiros (UFF, Niterói/RJ, Brasil)
Vera Lucia Lopes Cristovao (UEL, Londrina/PR, Brasil)
Vera Menezes (UFMG, Belo Horizonte/MG, Brasil)
Vilson José Leffa (UCPel, Pelotas/RS, Brasil)

Sumário / Contents

Apresentação / Introduction

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*

Plínio A. Barbosa

Tommaso Raso 1361

A segmentação da fala espontânea: aspectos prosódicos, funcionais e
aplicações para a tecnologia

*Spontaneous Speech Segmentation: Functional and Prosodic Aspects
with Applications for Automatic Segmentation*

Plínio A. Barbosa

Tommaso Raso 1397

FormantPro As a Tool for Speech Analysis and Segmentation

FormantPro como uma ferramenta para a análise e segmentação da fala

Yi Xu

Hong Gao 1435

Acoustic Models for the Automatic Identification of Prosodic
Boundaries in Spontaneous Speech

*Modelos acústicos para a identificação automática de fronteiras
prosódicas na fala espontânea*

Bárbara Helohá Falcão Teixeira

Maryualê Malvessi Mittmann 1455

Automatic Segmentation of Spontaneous Speech

Segmentação automática da fala espontânea

Brigitte Bigi

Christine Meunier 1489

Acoustic Correlates of Prosodic Boundaries in French: A Review of Corpus Data <i>Correlatos acústicos de fronteiras prosódicas em francês: uma revisão de dados de corpora</i>	
George Christodoulides	1531
Automatic Speech Segmentation in French <i>Segmentação automática da fala em francês</i>	
Philippe Martin	1551
Prosodic Segmentation and Grammatical Relations: The Direct Object in Kabyle (Berber) <i>Segmentação prosódica e relações gramaticais: o objeto direto em kabile (berbere)</i>	
Amina Mettouchi	1571
Prosody and Processing: Comprehension and Production of Topic-Comment and Subject-Predicate Structures in Brazilian Portuguese <i>Prosódia e processamento: compreensão e produção de estruturas de tópico e de sujeito no português brasileiro</i>	
Andressa Christine Oliveira da Silva Aline Alves Fonseca	1601
Complex Illocutive units in L-AcT: An Analysis of Non-Terminal Prosodic Breaks of Bound and Multiple Comments <i>Unidades Ilocucionárias Complexas na L-AcT: uma análise de quebras prosódicas não-terminais em Comentários Ligados e Comentários Múltiplos</i>	
Alessandro Panunzi Valentina Saccone	1647
Syntax, Prosody, Discourse and Information Structure: The Case for Unipartite Clauses. A View from Spoken Israeli Hebrew <i>Sintaxe, prosódia, discurso e estrutura informacional: o caso das orações unipartidas. Uma visão do hebraico falado em Israel</i>	
Shlomo Izre'el	1675



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

Plínio A. Barbosa

Universidade Estadual de Campinas, Campinas, São Paulo / Brazil

pabarbosa.unicampbr@gmail.com

Tommaso Raso

Universidade Federal de Minas Gerais, Belo Horizonte. Minas Gerais /Brazil

tommaso.raso@gmail.com

This issue of *Revista de Estudos da Linguagem* is dedicated to a theme addressed in several other initiatives promoted by its guest editors, along with colleagues from the international community. The theme, which in recent years has played an increasingly important role in the disciplines that study speech production and perception, is the segmentation of speech into smaller units addressed from both formal and functional perspectives, fundamentally under a theoretical approach coupled with an empirical focus. Among the main initiatives, we mention:

- Two international workshops (IV Leel and X Lablita International workshop *Unit of Reference for Spontaneous Speech and their Correlation Across Language*, held in August 2015 at UFMG; and the workshop *Spoken Corpora advances: prosody as the crux of speech segmentation, annotation and multilevel linguistic studies*, organized in June 2018 at Cape Town as part of the 20th International Congress of Linguists activities);

- The book *In Search for a Reference Unit of Spoken Language: A Corpus Driven Approach*, to be released soon by John Benjamins;
- A special issue of the Journal of Speech Sciences scheduled to come out in mid-2019.

All of these initiatives are dedicated to the prosodic segmentation of speech, a subject that has become increasingly central to understanding speech structuring at various levels, as well as the relationship of this structuring with the communicative functions of language. The disciplines interested in the subject, and Linguistics *in primis*, have evolved enormously from the contribution of technological advances and statistics applied to linguistic studies, and from the contribution of the advances of linguistic theories themselves. In fact, until recently, the study of speech segmentation considered almost exclusively the segmentation of the so-called lab speech. This includes read speech and speech elicited in various forms (XU, 2010) from the manipulation of external events (such as by proposing tasks with one or more participants such as map task and electronic games, conducting interviews on specific topics, *inter alia*). A few years ago, however, it became possible to approach good-quality, recorded non-scripted speech extracted from spontaneous speech corpora in varied natural communicative situations. In this introductory article to this thematic issue of RELIN, we present a partial overview of the scientific issues at stake, the results achieved so far, and the steps already announced for the future.

1. Prosodic segmentation: between form and function

Contrary to writing, which is a product that can be preserved in time and space, speech is a process whose result disappears shortly after its manifestation, if we set aside in this examination the current recording technologies. Only some cognitive consequences of discourse remain, but not speech itself (LINELL, 2005; BLANCHE-BENVENISTE; JEANJEAN, 1987). Absent from writing in its acoustic manifestation, except for mere indications inferred from punctuation marks, prosody is the essential component for speech segmentation studies. It is now possible, thanks to technology and dedicated software, to reproduce speech for as many times as necessary and to annotate the speech chain into different units by procedures of labelling and segmentation: syllables,

groups of syllables or words, prosodic units of different dimensions and theoretical status, as well as utterance sequences. This allows the systematic observation and measurement of many aspects of speech that, without technology, had to some extent only been intuited through the auditory sensitivity of the precursors of contemporary prosody research (see PIKE, 1945; LIEBERMAN, 1960; BOLINGER, 1965) without the possibility of being deepened or demonstrated. Among these aspects, a place of crucial importance is occupied by the different units in which it is possible to segment the flow of speech and by the development of a current of thought on its forms and functions. Finally, it has become possible to attempt the reconstruction of the complex prosodic structure (and not only) of human speech.

In addition, technology has made it possible to compile and investigate large amounts of speech data, treated and annotated in different ways and specifically suited to several research fronts, in a line with the view that privileges the acquisition of knowledge from huge corpora (cf. the concept of “big data” in FURHT; VILLANUSTRE, 2016). The automatic processing of the acoustic signal allows us to segment discourse into smaller units, from the utterance (or perhaps from larger units like the “paragraphs”) to the syllable and its constituents; furthermore, it allows us to investigate how human speech conveys boundaries (or their absence) at different hierarchical levels.

Depending on the interest of the study, the speech chain can be segmented into units of different sizes and types, conveying their own properties and delimited by some type of boundary. For the sake of exemplification, let us only look at the units above the word level. We can divide the speech chain into stress groups (or n-ary feet, groups of syllables up to a stressed syllable, in the case of right-hand languages), into prosodic units called intonational or tonal or prosodic groups, in sentences, or, under a syntactic perspective, in intonational phrases (IP), intermediate phrases (ip) and sentences. Each type of segmentation is directly or indirectly associated with a theoretical view, but in many cases this does not preclude an empirical investigation whose results can be analyzed in the light of different theoretical perspectives. In recent years, several corpora with prosodic annotation of the boundary have been compiled for different languages (AURAN *et al.*, 2004; DU BOIS *et al.*, 2000-2005; OSTENDORF *et al.*, 1996; CRESTI; MONEGLIA, 2005; SCHUURMAN *et al.*, 2003; IZRE’EL, 2002; RASO; MELLO, 2012; Forthcoming; METTOUCHI *et al.*, 2010; GAROFOLO *et al.*, 1993).

Any kind of segmentation implies the presence of a boundary, either actually perceived or theoretically proposed. Thus, the boundary can be understood as a physically perceived rupture, it may refer to a testable limit for the realization of linguistic phenomena, and it may further be considered as a region between two units, a region that can be auditorily perceived or not.

This thematic number seeks to study the segmentation of what can be considered as the reference unit of the speech process (IZRE'EL *et al.*, Forthcoming). The very notion of reference unit can be understood in different ways, but we can provisionally define it as a minimal unit of complete and autonomous communicative meaning that composes a spoken text (CRESTI, 2000; MONEGLIA; RASO, 2014). This definition can be challenged, but it allows us to have a point of departure.

All the aforementioned types of units, regardless of how they are defined, are separated by boundaries that are defined by highlighting greater or lesser perceptual or theoretical grounds, since hardly one of these two criteria completely excludes the other. In the articles in this thematic issue, a perceptual basis is always present, but some papers assign a greater weight to theoretical aspects, and these aspects vary from one article to another. With these differences of perspective, the concept of boundary changes as well.

Of theoretical nature are the boundaries of constituents in syntactic and informational approaches. This does not mean that they cannot be associated with prosodic boundaries, which constitute the primary interest of this thematic number. In fact, we understand that prosody guides syntactic interpretation, as in cases such as the sentence *A ovelha de raça brasileira* (The sheep of Brazilian breed; word-by-word: The-sheep-of-race-Brazilian). From this unit of writing, two utterances can be uttered in two distinct forms of grouping, where “/” represents a strong non-terminal boundary:

[A ovelha de raça] / [brasileira] vs. [A ovelha] / [de raça brasileira]

In the first case, it is a sheep born in Brazil from a non-informed breed and, in the second case, a sheep from a breed developed in Brazil. It is precisely the prosodic constituents that allow the proper scrutiny of the syntactic structure of each utterance. That is, prosody allows for disambiguation between the two possible interpretations, since the limited resources of writing do not allow deciding between the two possible

interpretations. In this example, the appropriate prosodic structure guides a single syntactic interpretation with syntactic and prosodic constituents being congruent, that is, having the same limits. Because of the prevalence of prosody, the authors of this thematic number who deal directly with the issue of speech segmentation take prosodic constituents as the only appropriate units related to the speech chain.

Furthermore, almost all contributions of this issue assume the organization of speech in units that can be considered extensive to intonational units. When we use the expression “intonational unit” in this panorama, however, we mean not only a unit organized by patterns of fundamental frequency (f_0), but also by patterns of duration and possibly voice quality. A single work (that of Ph. Martin) segments speech into accent phrases, which does not exclude the fact that a single accent phrase or a set of accent phrases coincide with an intonational unit. The segmentation in accent phrases can, therefore, be seen as an opportunity to investigate the internal structure of the intonational unit, thus enriching, and not contradicting, the perspectives that prefer to focus on the analysis of the intonational unit.

It is difficult to define the intonational unit without reference to perception or to a postulate of a theoretical nature. In general, the intonational unit is defined as a group of words (it can also be a single word and, in rare cases, where the emphasis on syllables comes into play, less than a word. In the latter case, the boundary is a perceptual consequence of the prominence of the unit) delimited between a prosodic boundary and the immediate subsequent boundary. The unit is characterized by a coherent f_0 contour separated both physically and perceptually from the preceding and following contours (DU BOIS *et al.*, 1992, p. 17; CRUTTENDEN, 1997). This definition masks some difficulties in capturing the properties of an intonational unit without reference to its boundaries, and, on the other hand, without identifying the boundary independently of the concept of intonational unit, there is a clear risk of circularity. The very definition of “coherent contour” is not completely satisfactory since we do not know clearly which parameters favour or break coherence.

From a functional point of view, the intonational unit can be studied and linguistically defined based on different perspectives. The main ones are the syntactic perspective, the informational perspective (CHAFE, 1994; RASO; MELLO, 2014) and the conversational

perspective (BARTH-WEINGARTEN, 2016). However, the very individualization of the intonational unit is problematic. In fact, the recognition of a coherent prosodic profile or a prosodic boundary is not always obvious. As regards the identification of a boundary, studies are usually based on the statistical agreement between annotators. In this kind of task, a certain chunk of speech is segmented into smaller units by a set of annotators. The agreement between them is used to identify a particular kind of boundary. Other approaches consider the perception of a boundary as associated to a particular f_0 movement visible by using a dedicated software, such as the so-called boundary tone, a movement of f_0 aligned to the end of the unit, in the framework of the Autosegmental-Metrical Theory (LADD, 1996; PIERREHUMBERT, 1980).

Statistical tests of inter-rater reliability show that the agreement among annotators for the identification of boundaries, and consequently of units, is very high (more than 80%, especially in the case of the terminal boundaries; MELLO *et al.*, 2012; MONEGLIA *et al.*, 2005; YOON *et al.*, 2004; BUHMANN *et al.*, 2002). It is therefore consensual that the intonational unit constitutes an important level of speech organization, although the reasons for this organization remain controversial. According to some authors, this segmentation of the speech chain is due to the limits of memory (cf. COWAN, 1998), which impose groupings of a limited number of syllables for linguistic processing. According to others, the units would have cognitive motivations (CHAFE, 1994; CROFT, 1995; BYBEE, 2010). As for yet a third view, the segmentation corresponds to units of a syntactic nature and therefore prosodic boundaries and syntactic boundaries would be correlated, especially in the phonological approaches of prosody that presuppose a mapping between syntactic constituents and the limits of prosodic units (NESPOR; VOGEL, 1986; SELKIRK, 1995). A fourth proposal, dominant in this thematic issue, attributes to the prosodic boundary the role of delimiting units of informational nature, independently of its syntactic organization. Others still see a correspondence between prosody and units of another discursive domain (COUPER-KUHLEN, 2004; SCHEGLOFF, 1998). Those who study prosody as correlated to linguistic domains of a non-syntactic nature also tend to consider prosody as a structural element implemented before the segmental elements (see the Frame/Content theory by MacNEILAGE, 1998). An interesting view within prosodic studies (HIRST; DI CRISTO, 1998; BARBOSA, 2006) attempts a

compromise between syntactic and prosodic constituents by proposing that the syntactic structure imposes some restrictions, but would not determine the position of the realized boundaries. In this proposal, the prosodic boundaries would only appear in positions compatible with the syntactic structuring without necessarily establishing constituents of this nature. After all, given a certain sentence, there are several positions compatible with the syntactic structuring where a boundary could be placed, with each position signalling a different cognitive-informational interpretation. On the other hand, many syntacticians have realized how prosody is essential for explaining particular structures that resist simple explanations in the framework of traditional syntactic theories. This is the case for the so-called insubordination phenomenon (EVANS; WATANABE, 2016, BOSSAGLIA *et al.*, Forthcoming). In such cases, the interpretability of the structure depends decisively on its prosodic coding.

2. The main theoretical questions

Previous research has also shown that the study of prosodic boundaries depends on speaking style and partially on the typology of the spoken text as well. In fact, until recently, research had focused on the study of prosodic segmentation in read texts or limited sequences performed in laboratory with interesting results, but that does not seem to be comparable with what happens in spontaneous speech, a priority objective of this issue. In prosody studies linked to syntax and phonology, laboratory speech is often used to test relations between prosody and syntax (as in the case of disambiguation and in the investigation of the relation between prosodic and phonological/syntactic constituents delimited by theoretical boundaries). Read texts present a much smaller number of variables than spontaneous speech, in addition to greater predictability (PRICE *et al.*, 1991). What is more, read speech is the sonorous realization of a written text, therefore being structured based on principles distinct from those of spontaneous speech.

Recently, some works on spontaneous speech have obtained promising results in the investigation of segmentation mechanisms. This has been done either by observing a high agreement (greater than 80%) among human annotators (MELLO *et al.*, 2012; MONEGLIA *et al.*, 2005; TEIXEIRA FALCÃO, 2017) or by developing software able to segment spontaneous speech automatically, achieving results that are

highly comparable with the tasks performed by humans (AVANZI *et al.*, 2008; NI *et al.*, 2012; MITTMAN; BARBOSA, 2016).

The development of software capable of automating prosodic segmentation in intonation units (cf. MITTMAN; BARBOSA, 2016) is only possible because the investigation of the acoustic parameters responsible for boundary perception has greatly advanced, thanks to the work done with read speech and speech sequences performed in the laboratory, which allowed a first understanding of the highly complex phenomena at play. From that, it came up that the parameters responsible for our perception boundaries are diverse; they are not always all co-present; their weight may vary depending on the languages and the circumstances of a particular speech style. This leads to the question of whether it is possible to speak of boundaries as a homogenous category at all, and points in the direction in favour of speaking of different types of boundaries.

In the literature, the parameters that are most mentioned as boundaries markers are fundamental frequency (f_0), duration and intensity, as well as parameters that refer to voice quality (BARTH-WEINGARTEN, 2016; MO *et al.*, 2008; WAGNER; WATSON, 2010), especially creaky voice (DILLEY *et al.* 1996; GORDON; LADEFOGED. 2001; REDI; SHATTUCK-HUFNAGEL, 2001; HANSON *et al.*, 2001; CARLSON *et al.*, 2005). From them, the main boundary cues that emerge are: the silent pause, which we will simply call “pause” (later on we will discuss the role of the filled pause), whose presence automatically seems to convey the perception of a boundary (MARTIN, 1973; SWERTS, 1997; SHRIBERG *et al.* 2000; TSENG; CHANG 2008; MO; COLE 2010; TYLER, 2013); the lengthening of the final syllables of the unit, that is, a decreasing of speech rate during the last syllables before a boundary (WIGHTMAN *et al.*, 1992; BARBOSA, 2008; MO *et al.*, 2008; FUCHS *et al.*, 2010; FON *et al.*, 2011; TYLER, 2013); the shortening of the first syllables of the unit, that is, speech rate increases just after a boundary (AMIR *et al.* 2004; TYLER, 2013), correlated with phenomena of anacrusis; the reset of the f_0 curve; the abrupt change of direction of the f_0 curve; the change of intensity at the beginning of the prosodic unit (SWERTS *et al.*, 1994; TSENG; FU, 2005; MO, 2008); creaky voice and perhaps other non-modal voice qualities. To these parameters, at least for some languages, some phenomena of a segmental nature must be added. For example, for English, final stop release and creakiness or glottal closure in the vicinity of final segments may be cues of a boundary.

Each of these cues brings some issues for the researcher. For example, the pause, which intuitively seems an obvious notion, is not identified consensually: what is the minimum amount of silence considered as a pause? How does the presence of a pause affect the other parameters that contribute to boundary perception? Is the pause a clue of boundary type or not? As for the f_0 curve, what is the relative contribution of f_0 level difference, f_0 excursion, the direction of f_0 movement, and of f_0 variation rate? When considering syllabic duration, what is the extent of the region affected by the boundary, measured in number of syllables? Additionally, if the change in duration involves more than the syllable just before and after the boundary, does the change occur in the same proportion for each syllable involved or not? Furthermore, previous experimental work has shown that, in order to reliably evaluate duration measures, some form of normalization that sets aside the intrinsic properties of the segments is necessary, which, in this case, decisively influences the duration (BARBOSA, 2012). It should also be noted that the measure of duration appropriate for prosodic analysis should consider phonological and phonetic syllables. The former is important for the perception of speech, because it involves syllable perception through the cognitive system, while the latter is the basis for the production of the speech chain and the structural organization of the corresponding consonants and vowels.

Research on the acoustic parameters that, together, convey the perception of a boundary should consider the weight or relative contribution of each acoustic cue. For this, it is important to consider not only that each cue is perceptible only if it surpasses a certain threshold, but that this threshold varies by varying the other cues (t'HART *et al.*, 1990). This means, first, that we are not able to perceive just any change in f_0 or any change in duration or intensity, but only changes that exceed a certain threshold. Although for each parameter or cue in isolation we can know its *Just Noticeable Difference* (JND), that is, the minimum variation of this parameter that we can perceive (see HUGGINS, 1972; KLATT; COOPER, 1975 for segmental duration, t'HART, 1981; and RIETVELD; GUSSENHOVEN, 1985, for f_0 as well as KOFFI, 2018, for intensity), as well as the way in which the JND varies with the modification of another parameter (for example how we perceive intensity variation at different frequencies), we do not know yet how these complex combinations of parameters vary with respect to the ability to convey boundary perception.

It is not simple to model the boundary phenomenon given the possibility of combining so many parameters in the speech flow. In fact, it would not be surprising if the weight of a cue changes by changing the combinations of the other cues, or by changing speaking style - reading or spontaneous conversation, or other styles of spontaneous speech, or different linguistic functions of the units delimited by the boundaries, without considering variations related to the characteristics of the speakers.

In fact, the studies in different languages confirm the importance of the aforementioned cues for the perception of a boundary, while revealing that each one of these cues acts with a distinct weight to mark this same boundary (TEIXEIRA FALCÃO, 2017). This varying hierarchy of acoustic cues seems to be linked to the functions that a certain parameter has in the language. For example, in tonal languages, f_0 has the role of conveying linguistic functions that in non-tonal languages are conveyed by other parameters. In these languages, f_0 differences implement tone distinctions that serve to contrast lexical items. In addition, the weight of f_0 is affected when this parameter is used to mark the boundary, with duration and f_0 reset being the most relevant parameters for signalling boundaries (YANG; WANG, 2002). This is likely to be the case with other parameters, which would behave differently to signal the prosodic boundary depending on how important they are to convey other functions in a given language. Very little is known about how the weight of a given parameter changes within a large combination of other parameters for marking boundaries of functionally different units.

While some studies focus on investigating the opposition between presence vs. absence of a boundary (MO *et al.*, 2008; BARBOSA, 2010), other studies investigate a potential diversity among the boundaries. In the latter case, some authors propose the existence of a certain number of boundaries, while others propose a continuum between presence and absence of boundaries. In this second case, there is always a risk of finding some degree of boundary, no matter how small, and losing the boundary vs. non-boundary contrast, making any consideration of a functional nature attributable to a boundary extremely difficult, if not impossible.

On the other hand, the researchers who consider that the boundary is a gradient phenomenon, although categorical, propose a gradation of strength for the different boundaries, which occur in a limited number. Among these authors there is disagreement about the amount of different strengths that can be recognized and perceived (see BARBOSA, 2006,

for a discussion). Some studies distinguish between strong and weak boundaries, while others consider it possible to individualize more than two degrees of strength (see WIGHTMAN *et al.*, 1992, for English, BARBOSA, 2006, for Brazilian Portuguese, and BARBOSA, 1994, for French) with some of them reaching up to seven degrees, which is in line with the phonological theories for prosody such as those by Nespor and Vogel (1986) and Selkirk (1995).

Another possibility to infer degrees of boundary strength is the use of local maxima of the acoustic parameters that convey a prosodic boundary as indices of the strength of this boundary (TEIXEIRA FALCÃO, 2017). Even if local maxima vary continuously, it is possible to use clustering techniques to infer a limited number of boundary strengths that do not exceed four (see BARBOSA, 2006, for BP, and BARBOSA, 1994, for French). In the work for BP, Barbosa (2006) used z-score-normalized syllable duration maxima to obtain 3 to 4 distinct levels, partially correlated with syntactic boundaries obtained by the projection of a dependency tree in line with Tesnière's (1965). The different degrees of strength allowed establishing a hierarchy of prosodic constituents that open the possibility of inferring the prosodic structure of an utterance. This procedure had already been proposed by Grosjean and colleagues (GROSJEAN; GROSJEAN; LANE, 1979; GROSJEAN; DOMMERGUES, 1983; GEE; GROSJEAN, 1983) by asking people to read at increasingly slow rates and subsequently analysing vowel durations associated with silent pauses when applicable and from segmentation indices for utterances obtained from perception tests. This procedure reveals what they called a *structure de performance*, a prosodic structure with the following properties: constituents of similar size, hierarchical organization and symmetric structure (GROSJEAN; DOMMERGUES, 1983). These properties emerged from two competing constraints: the speaker's tendency to respect the linguistic structure of the sentence and the tendency to balance the extension of the constituents it produces (MONNIN; GROSJEAN, 1993, p. 28; MARTIN, 1987). The tendency to equilibrate the extension of prosodic constituents would explain why subjects do not systematically group the verb with the object noun phrase when pronouncing English phrases, as would be predicted by syntax, but prefer groupings of type (SV)O (GROSJEAN; GROSJEAN; LANE 1979, p. 59).

The discussion about boundary types, however, is not just quantitative in nature. Many authors distinguish between boundaries that convey perception of prosodic and linguistic completion (with distinct interpretations of the nature of the completed linguistic unit) and boundaries that convey the perception of discourse continuity. The latter signals that the discourse segment in progress cannot be considered complete even if the boundary signals the end of a constituent, this one having distinct types, depending on the theoretical approach (MONEGLIA; CRESTI 1997; CRYSTAL, 1969; SWERTS, 1994; SWERTS *et al.*, 1994). For several authors these two types of boundaries are called terminal and nonterminal, respectively.

But some authors who consider the distinction between terminal and non-terminal boundaries argue for a fine-grained difference. For these authors, there would not be a single type of terminal boundary, nor a single type of non-terminal boundary. According to this proposal, we can observe that some terminal boundaries are “more terminal” than others. For example, the boundaries of utterances would be less terminal compared to the boundary between larger discursive blocks, called paragraphs by some (van DONZEL 1999). Similarly, there would be several types of non-terminal boundaries, some more prominent than others, or perceptually closer to the terminal boundaries, or announcing the fact that the conclusion is close. These proposals should not be considered as mutually exclusive, since they are able of capturing different aspects of the complexity of the phenomenon (SWERTS *et al.*, 1994; TEIXEIRA FALCÃO 2017).

In fact, if we examine the phonetic-acoustic parameters correlated to boundary perception, in particular the non-terminal boundary, we observe varied combinations within the same language and text (see TEIXEIRA FALCÃO, 2017). We have, for example, boundaries clearly marked by a movement of increasing f_0 , an acoustic cue of continuity, which, along with other prosodic cues like duration, conveys the perception that the discourse will continue. On the other hand, this increasing movement of f_0 or final lengthening may be lacking in other boundaries that are also perceived as non-terminal (WAGNER, 2010).

As for conclusive boundaries, it is often observed that they are characterized by a downward movement of f_0 to the lowest level, followed by a reset of f_0 at the beginning of the next unit, which would start with an f_0 value at a clearly-defined distinct height. However, it

is commonly recognized that not all utterances conclude with a low f_0 value. Although the most obvious and studied case is that of the yes/no questions in languages such as English and Peninsular Spanish, there are other illocutions, according to the terminology and categorization we adopt, which are marked, among other parameters, by a higher f_0 at the end (CRESTI 2000; Forthcoming; MORAES; RILLIARD, 2014 *inter alia*).

The variability of the physical realization of the boundaries can be correlated with different functional values on the linguistic plane. We would then have not only a correlation between types of boundaries conveying completion and types of boundaries conveying continuation, but also between different conclusive types, in the case of different illocutions, and between different non-conclusive types, which, by hypothesis, would mark different constituent types (syntactic or other kinds). In this perspective, the specific realization of a prosodic boundary would not only have a demarcating value, but would depend heavily on the linguistic function of the unit delimited by the boundaries, the associated cues would also point to these same linguistic functions.

Thus, in this perspective, studying how boundaries are physically realized and studying the nature of the units delimited by these very boundaries (one on the left and the other on the right) would no longer belong to distinct scopes. The former having been of a prior interest to Phonetics and the latter to those who are interested in higher linguistic levels or in cognitive mechanisms would, therefore, become much more integrated. The perspective that unites the functions of the units to the concrete manifestation of the boundaries that delimit them is still incipient and can give us interesting answers about the nature of the units that are delimited by these boundaries.

Before moving on to the different theoretical approaches to units, it is worth making an observation about some kinds of boundaries (and units) that are much less frequent in laboratory speech, at least in the case of read speech, but which are extremely common in spontaneous speech: the different types of disfluencies. In spontaneous speech, the phenomena of interruption, retractings and hesitation are very frequent. Many units come to an end not because the speaker planned their completion, but because some unforeseen internal (improper word retrieval, change of mind, or any problem in the articulation or elaboration of content) or external cause (interruption by another speaker or any environmental

event) leads to the momentary interruption of the utterance before it is completed semantically and prosodically. As for retraction, the statement is not interrupted, but is fragmented by repetitions of words or parts of words, which the speaker then ideally cancels and corrects, continuing to produce the utterance as if they had not been pronounced. This is the result of difficulties in the realization of the utterance that do not lead to interruption of the statement and are more or less present in all speakers, but especially in those who have less mastery of speech, or because they are very young, or because they are from a lower diastrophic category, or for other reasons. In the case of hesitation, difficulties in speech are manifested under different guises, such as vowel stretching or time taking by producing filled pauses (e.g., anh, ehh). One or two boundaries (one in the case of the interruption and usually two in the other two cases) always or nearly always occur when one of these three phenomena takes place. However, in principle, these boundaries are not planned by the speaker and do not mark units with a linguistic function. In the analysis of the prosodic boundary cues, they constitute an element of noise, and cannot be compared to the boundaries that the speaker makes to build the meaning of the utterance.

A last type of boundary we have to consider is the one that delimits the units that, in the model of the *Language into Act Theory* (L-AcT; CRESTI, 2000; MONEGLIA; RASO, 2014; MONEGLIA; CRESTI, 1997), are called *Scanning Units*. A *Scanning Unit*, according to L-AcT, is an informationally non-autonomous unit constituting one part of a bigger information unit (e.g. a Topic divided into two or more intonation units). In this case, the units before the last one are *Scanning Units*, and the prosodic profile conveying the information unit function always appears in the last intonation unit. For L-AcT, boundaries that delimit these types of units are due to different possible reasons: emphasis (in order to make parts of an information unit text prominent, its content is segmented into more intonation units); lack of skill in speech (such as small hesitations or retractings without any added segmental material); articulatory necessity (when an information unit features too many syllables for them to fit comfortably in one intonation unit). These kinds of boundaries that, as we have seen, do not constitute a homogeneous group, constitute a problematic typology with regard to the other kinds of boundaries, since the individualization of a *Scanning Unit* is possible

only after a text has been informationally annotated, and this annotation follows text segmentation and cannot be automatized.

Besides these open issues, it would also be interesting to consider some other non-linguistic ones: do male and female voices use the acoustic parameters that convey perception of boundary in the same way? What happens in the different speech pathologies, in which articulatory or cognitive functions are endangered? How do skills that deal with this functional goal develop along ontogenesis?

Along the past decades, research has greatly improved its investigation and understanding of the complex combinations of factors that affect boundary expression; more recently many works have begun the investigation of this phenomenon in spontaneous speech. However, there remains a long way to be covered. Finally, to face the parameter problem is still not sufficient. It is necessary also to look carefully at each parameter in their different combinations and at their weight (hierarchy) in each combination. Of course, this increases the variables responsible for signaling prosodic boundaries, and imposes the use of computational and statistical tools in order for them to be satisfactorily captured.

More recently, prosodic boundaries have been the object of psycholinguistic investigations in an attempt to better understand how their perception is processed (DRURY *et al.*, 2016; GLUSHKO, *et al.*, 2016; NICKELS *et al.*, 2013; HWANG; STEINHAUER, 2011; PAUKER *et al.*, 2011; STEINHAUSER, 2003; STEINHAUER; FRIEDERICI, 2001), especially through the *Event-Related Potential* (ERP) technique. Steinhauser *et al.* (1999) were the first ones who used this technique to show that perceived prosodic boundaries are associated to intervals of increased amplitude in electric activity (evoked potential), named CPS (*Closure Positive Shift*). This peak occurs between 400 and 800 ms. after a defined moment, which, in the most successful tests, was considered in the last stressed syllable before the boundary. The experiments were performed with and without the presence of pause and of other parameters considered responsible for conveying the perception of boundary, but the electric activity peak was always detected. It seems that syllabic lengthening and the presence of a boundary tone are sufficient to trigger the hearer's encephalic reaction. Currently, researchers are trying to refine further the observation of human reaction to isolated parameters, or to their combinations, for the perception of boundaries.

The fact that segmentation (*phrasing*) seems to be sensible to cues of different modalities is especially interesting: not only acoustic cues, but also graphic ones, such as commas in reading, seem to cause an increase of electric activity when there is a boundary. Besides this, the phenomenon also occurs for musical segmentation, but with a greater latency (may be due to the lack of linguistic information, like syntax or lexicon). It also seems that CPS can be encountered only after a certain age (more or less three years of age), and this could be explained if we consider that it depends on a minimal capacity for structuring, either syntactically or prosodically, *stricto sensu*. This result is compatible with data about language acquisition (THORNTON, 2016; HYAMS; ORFITELLI, 2015 *inter alia*). Finally, CPS seems to be more evident when the boundary is less expected, that is, when it is not or is minimally predictable based on information of different natures; but it also seems clear that prosody, as a vehicle for boundaries, prevails when it is in conflict with syntactic expectations (BÖGELS; TORREIRA, 2015; BÖGELS *et al.*, 2013; 2010).

Because boundaries are marked by the combination of all the prosodic parameters, mainly syllabic duration, f0 and intensity, it is important to add that dextral individuals have a predominant temporal processing in the left hemisphere, while spectral processes mainly activate areas of the right hemisphere (ROBIN *et al.*, 1990; ZATORRE, 1997). This is confirmed by studies on impaired individuals, either on the left or on the right hemisphere, the former losing capacity of temporal processing (SHAH *et al.*, 2006). As far as the neuronal areas involved in speech perception, both temporal cortical areas and parietal ones are bilaterally activated (HICKOK; POEPPPEL, 2000).

3. Segmentation and linguistic meaning

Speech segmentation is essential to build linguistic meaning (cf. FERY, 2017, for a review). Prosody is used to mentor the hearer in reconstructing the different functional units and their hierarchy and function, in order to decode the message. This is the main reason that motivates researchers to study the physical nature of boundaries and its relation with the different linguistic levels. Let us look at some examples in different languages.

In English, a sequence as *People give John the book I promised him* can be parsed at least in the four following ways, giving rise to very different meanings, from both illocutive and syntactic points of view:

- (a) *People* (Calling)! *Give John the book I promised him* (Order)!
- (b) *People give John the book I promised him* (Assertion).
- (c) *People give John the book* (Question)? *I promised him* (Assertion).
- (d) *People* (Calling)! *Give John the book* (Order)! *I promised him* (Assertion).

In (a), (c) and (d) we find two terminal boundaries, while in (b) we find just one, which is terminal, too. However, when we look at the acoustic parameters, terminal boundaries associated to the different possible segmentations vary, at least as far as f₀ movements are concerned. If the second boundary in (a), (c) and (d) is preceded by a falling movement, the first boundary features a rising one. These rising movements are different, as much as the different falling movements of the other cases. A similar distinction could be made for the values of duration and intensity.

In Portuguese, a sequence such as *João vai pro Rio até amanhã* (*João will go* (or *go*) *to Rio until tomorrow* (or *see you tomorrow*) can be parsed at least in three different ways:

- a) *João* (calling)! *Vai pro Rio até amanhã* (order)! (*João! Go to Rio until tomorrow*)
- b) *João vai pro Rio até amanhã* (assertion). (*João will go to Rio until tomorrow*)
- c) *João* (calling)! *Vai pro Rio* (order)! *Até amanhã* (greeting)! (*João! Go to Rio! See you tomorrow*)

In these three sentential organizations, it is evident that segmentation affects the syntactic and the semantic-pragmatic interpretation of the sequence.

Finally, the following example shows how segmentation can decide syntactic and semantic interpretation in Italian:

- (a) *Claudia* (calling)! *Guarda* (deixis)! *Quanto è bello* (expressive)!
(*Claudia! Look! How beautiful it is!*)
- (b) *Claudia* (calling)! *Guarda quanto è bello* (deixis)! (*Claudia! Look how beautiful it is!*)
- (c) *Claudia guarda quanto è bello* (assertion). (*Claudia looks how beautiful it is.*)

The series of examples could easily be more complex, considering different interpretations and other types of units. It could also easily be extended to other languages. However, what is relevant for us is to make the importance of the role of prosodic parsing in the construction of linguistic meaning evident, both at the syntactic and at the semantic level. The presence of a boundary certainly affects the phono-morphological level too, for instance, inhibiting sandhi phenomena.

In the previous examples, we have observed some cases of terminal boundaries; they isolate pragmatically and prosodically autonomous linguistic sequences that can be uttered in isolation. However, meaning is also affected in the case of non-terminal boundaries, that is, when the (syntactic or informational) relationship between two units separated by a boundary must be maintained. For example, the sequence *the film I like it* can be analyzed as a noun phrase modified by a relative clause. However, if we insert a boundary, the analysis can change: in *the film, I like it* the analysis can show a Topic-Comment relationship that can be interpreted like: *as for the film* (TOP), *I like it* (COM).

Let us go back to the notion of *unit of reference for speech*, as the minimal unit of the text that carries an autonomous communicative (in the actional sense) meaning. If we consider the prosodic dimension, it is hard to define this unit only through the syntactic criteria used to characterize traditional categories like clause or sentence. Prosody has a communicative dimension that leads researchers to rather pay attention to production and perception of speech, even if we do not lack more abstract perspectives (but possible only outside a communicative context).

Many of the linguists who incorporate prosody as one of the main elements of their models consider prosodic perception of terminality of a communicative sequence as the main cue of the unit of reference for speech (CRESTI, 2000; MONEGLIA; RASO, 2014; IZRE'EL, 2002). Others prefer to consider the intonation unit as unit of reference, no

matter if its prosodic profile is perceived as conclusive or non-conclusive (METTOUCHI *et al.*, 2010). In both these perspectives, the main cue that defines a unit of reference corresponds to the boundary of an intonation unit. The difference consists on whether any kind of boundary determines a reference unit or only boundaries with a specific quality can do it. This discussion goes along with that concerned with the linguistic relations that occur within an intonation unit, those that occur among different intonation units pertaining to the same terminated sequence, and also those across the boundary between different terminated sequences (for some aspects of this discussion in a different but similar framework, see Izre'el in this volume; CRESTI, 2014; PIETRANDREA *et al.*, 2014).

4. The papers in this volume and their contribution to the debate

The nine papers presented in this thematic volume deal with different aspects of prosodic segmentation of spontaneous speech. A first group of papers focuses on the development of software that allow the extraction of data and information useful to clarify some of the many questions related to prosodic segmentation. Of course, also behind these works there is a theoretical hypothesis, either about the function or the number of different boundaries to be identified.

The paper by Xu and Gao presents the FormantPro script, which uses the software Praat as its platform for the automatic extraction of formant trajectories. Although the theme of this article does not directly focus on the problem of prosodic segmentation, the tool and the examples that the authors bring open a discussion about the isomorphism between acoustic and articulatory events that delimit the boundaries of consonants and vowels. These boundaries are discussed with relation to the issue of the alignment of these segmental landmarks with trajectories of f_0 that eventually might have implications to delimitate prosodic boundaries. The software also generates values of duration and intensity and allows the presentation of the mean trajectories in terms of temporal normalization, which helps observing the equivalencies among instances of different utterances with words in contrast. The values of duration can be used to investigate cues of prosodic boundaries in case of important changes with respect to context.

The work by Teixeira Falcão and Mittmann presents an interesting procedure to extract models of acoustic parameters for different types

of boundaries in stretches of spontaneous speech corpora previously segmented by 14 segmentators. The data from corpora were treated to make them readable by the script in Praat. After this, a very high number of measurements is extracted in a window of ten V-V units to the left and 10 V-V units to the right of each position that is a candidate to be a phonological word boundary. The V-V segmentation (BARBOSA, 2006) shows how other levels of speech segmentation necessarily interact with the level of the intonation unit. A statistical procedure, after human refinement, reveals the combinations of parameters that better explain the boundaries and their weight. The whole work was planned considering that prosodic boundaries can be distributed into two big groups: terminal and non-terminal. The work about non-terminal boundaries suggests that it would be necessary to consider these boundaries as at least three different sub-groups, with three different models to account for non-terminal boundaries. These findings encourage the hypothesis that we should differentiate between terminal and non-terminal boundaries, and that we need more subtle distinctions. It would be very important to investigate what accounts for the latter.

The paper by Bigi and Meunier evaluates the SPPAS software, which allows the automatic segmentation of read and spontaneous speech, placing main focus on disfluencies found in spontaneous speech. The tool presupposes the existence of an orthographic transcription and a lexicon pronunciation dictionary. It uses an acoustic model of the sounds of French speech, which allows the alignment of phonetic symbols with the speech signal. The errors in the alignment are approximately 11% in read speech and 15% in spontaneous speech, but they can be reduced using an enriched orthographic transcription that identifies disfluency types. The tool has been tested in nine corpora, including read speech, spontaneous conversation and political debates, for the cases with disfluencies, laughter, filled pauses and noises. The authors show that, when preceded by a pre-processing that segments the speech flow into inter-pausal units, it is possible to achieve a precision level of about 20 ms in the segmentation task.

The article by G. Christodoulides uses two French spoken corpora with the annotation of boundaries of different strengths, in order to verify: (a) degree of agreement between prosodic annotations originated from two different theoretical perspectives, the autosegmental-metrical theory (PIERREHUMBERT, 1980) and the distinction between micro

and macro-syntax (BLANCHE-BENVENISTE, 2002; 2003) referring to two comparable levels of annotation; (b) which acoustic parameters are more important to convey the two types of boundaries and what their hierarchy is. The use of corpora depending on such different theoretical perspectives is an important test for research about prosodic boundaries. This is even more true considering that one corpus is segmented based on theoretical criteria and the other based on perceptual ones. The investigated parameters are: presence and duration of pause, pre-boundary lengthening and two measurements of f_0 associated with boundary. The analysis shows a very high agreement between the two corpora as far as the prosodic parameters in the positions marked as boundary and the distinction between the two types of comparable boundaries are concerned. The conclusion is that the most important parameters associated with boundary and boundary strength is pause, followed by syllabic lengthening. f_0 seems to be important to distinguish between presence or absence of boundaries, but not to signal boundary strength and therefore distinguish the two types of boundaries.

Ph. Martin's work differs from the others because it analyses a different unit: the stress group. The object of the paper is therefore a smaller unit than intonation unit, even if sometimes the two units may coincide. Martin individualizes a limited number of possible f_0 movements in the stress group inside the intonation unit, and observes that there is a dependency criterion among them. This allows us to investigate the internal structure of an intonation unit, based on smaller units marked by stress. Among other consequences, the results of this analysis may bring to light some characteristics of the internal structures of different intonation units, and may show how these structures correlate with the linguistic function of a specific intonation unit. Different aspects of the unit, with the presence of some prominences in defined positions, have already been discussed in the literature, even if not conclusively in our perspective. Proposals like that by Martin lead us to consider the role played by other prosodic levels and their specific linguistic functions, that, besides other characteristics (prominences, type of boundary), may give us a better understanding of how we build a sequence with a definite linguistic function dealing with different levels of the prosodic structure.

A third group of papers investigates the boundaries clearly with linguistic goals, either syntactic or informational.

The study by A. Mettouchi on Kabyle, an Afro-Asiatic language of Algeria, shows how the presence/absence of a boundary can constitute the linguistic cue that marks a syntactic function, in this case the direct object. The boundary reveals itself as the decisive cue in order to distinguish this structure from structures that can have different functions, probably informational ones, but that appear in the utterance with the same formal cues, except for the presence (other functions) or absence (direct object) of a prosodic boundary. This study raises an important issue: the relationship between the presence of boundaries and the rupture of syntactic compositionality. Other studies (CRESTI, 2014; RASO; VIEIRA, 2016; BOSSAGLIA *et al.*, Forthcoming) treat this important aspect, which is still controversial. If on the one hand it is easy to find cases in which it seems clear that syntactic compositionality is interrupted where there is a prosodic boundary (making it possible to think that some type of boundary has the possibility of marking this interruption), on the other hand, we still have cases that are interpretable, thus saving the syntactic compositionality across a prosodic boundary.

The article by da Silva and Fonseca also presents several aspects of interest. The first one, as with the previous and the following studies, is the importance that a prosodic cue has for the identification of a linguistic unit, in this case the unit of Topic. The second reason is the experimental basis of the research, about which we will come back later. A third reason is that the work shows how results presented within a formalist framework can also be useful for the study of Topic in different perspectives, making it clear how the empirical view on data can benefit the scientific debate. The experiments idealized and implemented by da Silva and Fonseca can be of great interest for the debate among researchers about information structure in speech. The results can be used to compare a syntactic definition of Topic with definitions of a pragmatic nature, especially the one proposed by L-AcT, which assigns to prosody a crucial weight besides presenting many results investigating different languages, among which BP (cf. CRESTI, 2000; SIGNORINI 2004; FIRENZUOLI; SIGNORINI, 2003; MONEGLIA; RASO, 2014; ROCHA; RASO, 2013; CAVALCANTE, 2016; MITTMANN, 2012; RASO; CAVALCANTE; MITTMANN, Forthcoming). Actually, the non-expected results found for the third experiment reported in the paper could be easily explained assuming that Topic is a pragmatic category that does not depend on argument structure and, therefore, can

occupy the subject position, but is marked by a prosodic boundary and a functional prosodic focus that distinguish it from subject. The subject, on the contrary, does not present a prosodic boundary between itself and the rest of the utterance and does not carry any prosodic functional focus. In this case, the difference between Topic and subject would not consist in their being two different syntactic functions, but would be explained as a difference of linguistic level: the subject would be a syntactic function and an argument of the verb in the Comment unit, while the Topic would be a pragmatic function, external to the Comment unit. A more in-depth debate between these different theoretical perspectives could clarify the notion of Topic and stimulate both approaches to refine their analyses and their argumentation, using both experimental procedures, like those proposed by da Silva and Fonseca, and data extracted from spontaneous speech corpora, like those compiled taking L-AcT into account (CRESTI; MONEGLIA, 2005; RASO; MELLO, 2012; Forthcoming).

The study by Panunzi and Saccone is also clearly theory-oriented. In fact, its goal is to observe if, to which extent and how boundaries between different pairs of information units are performed in different ways. The two pairs (rarely sequences of more than two items) that are explored in the article are different combinations of illocutionary units. One type of pair is characterized by two prosodically and pragmatically patterned illocutions that build a unique interpretation. The other type, on the contrary, is constituted by two independent illocutions, even if separated by a non-terminal boundary. Therefore, in order to analyze the boundaries, the text must be informationally tagged according to a theoretical framework, in this case L-AcT (CRESTI, 2000; MONEGLIA; RASO, 2014). The first results suggest that there are clear formal differences between the two pairs of units. This is an intriguing example showing how characteristics of the boundary may correlate with the function of the units separated by it. This kind of study, which tries to correlate linguistic functions of the intonation unit and boundary cues, can be applied to different kinds of units and can be based on different theoretical frameworks.

The paper by Izre'el is the last one in this volume because, based on some considerations about the linguistic role played by prosody and especially by prosodic boundaries, it proposes a general revision of the traditional categories phrase, clause, sentence and predication, showing how the incorporation of prosody may lead to a general reformulation of

canonical categories in the study of spontaneous speech. Izre'el revisits the discussion about these categories starting with the ancient Greek tradition up to Chomsky, in order to show how some categories, as they are defined in the syntactic tradition, do not work in the analysis of speech, especially of spontaneous speech, which, in principle, should be the natural domain for the analysis of language. Considering prosody and data from spontaneous speech corpora, the importance of the illocution (which Izre'el calls *modality*) clearly emerges as a crucial category to individualize the communicative unit and as a prosodically marked category. The importance of prosodic boundaries also clearly emerges as a means to define the domain of linguistic relations in their communicative realization. Like other papers in this volume, but portraying a wider scope, this paper brings more arguments to the discussion (cf. also BIBER *et al.*, 1999; the papers in RASO; MELLO, 2014; CRESTI, 2005; RASO; MITTMANN, 2012, *inter alia*). It highlights the urgency of defining the communicative unit of speech, of revising the notion of predication (and of proposition), or those of clause and sentence, and sustains how important it is to incorporate prosody as the central element to mark the unit of reference for spoken communication. As other articles in this thematic issue, the paper by Izre'el does not leave any doubt about the necessity of incorporating prosody among the levels of linguistic analysis, and, more than this, about the crucial hierarchical weight of prosody to individualize the linguistic constituents of speech.

References

- AMIR, N.; SILBER-VAROD, V.; IZRE'EL, S. Characteristics of Intonation Unit Boundaries in Spontaneous Spoken Hebrew: Perception and Acoustic Correlates. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 2004. Nara. *Proceedings...* Nara: ISCA, 2004. p. 677-680.
- AURAN, C.; BOUZON, C.; HIRST, D. The Aix-MARSEC Project: an Evolutive Database of Spoken British English. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 2004. Nara, Japan. *Proceedings...* Nara: ISCA, 2004.
- AVANZI, M.; LACHERET-DUJOUR, A.; VICTORRI, B. ANALOR. Tool for Semi-Automatic Annotation of French Prosodic Structure. In: ANALOR. A Tool for Semi-Automatic Annotation of French Prosodic Structure. Campinas, Brazil, May 2008. p. 119-122.

BARBOSA, P. A. *Caractérisation et génération automatique de la structuration rythmique du français*. 1994. Tese (Doutorado) – Institut National Polytechnique de Grenoble, França, 1994.

BARBOSA, P. A. *Incurções em torno do ritmo da fala*. Campinas: Pontes, 2006.

BARBOSA, P. A. Prominence-and Boundary-Related Acoustic Correlations in Brazilian Portuguese Read and Spontaneous Speech. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 4., 2008, Campinas. *Proceedings...* Campinas: ISCA, 2008. p. 257-260.

BARBOSA, P. A. Automatic Duration-Related Saliency Detection in Brazilian Portuguese Read and Spontaneous Speech. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 5., 2010, Chicago. *Proceedings...* Chicago: ISCA, 2010.

BARBOSA, P. A. Panorama of Experimental Prosody Research. In: GSCP INTERNATIONAL CONFERENCE – SPEECH AND CORPORA, VIIth., Belo Horizonte. *Proceedings...* Florence: Firenze University Press, 2012. p. 33-42.

BARTH-WEINGARTEN, D. *Intonation Units Revisited*. Cesura in Talk-In-Interaction. Amsterdam: John Benjamins, 2016.

BIBER, D.; JOHANSSON, S.; LEECH, G.; CONRAD, S.; FINEGAN, E. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited, 1999.

BLANCHE-BENVENISTE, C. Macro-syntaxe et micro-syntaxe: les dispositifs de la rection verbale. In: ANDERSEN, H. L.; NØLKE, H. (Éd.). *Macro-Syntaxe e Macro-Sémantique*. Bern: Peter Lang, 2002. p. 95-115.

BLANCHE-BENVENISTE, C. Le recouvrement de la syntaxe et de la macro-syntaxe. In: SCARANO, A. (Ed.). *Macro-syntaxe et pragmatique*. L'analyse linguistique de l'oral. Roma: Bulzoni, 2003. p. 53-75.

BLANCHE-BENVENISTE, C.; JEANJEAN, C. *Le français parlé*. Transcription et édition. Paris: Didier Érudition; Institut National de la Langue Française, 1987.

BLANCHE-BENVENISTE C. Macro-syntaxe et micro-syntaxe: les dispositifs de la rection verbale. In: ANDERSEN, H. L.; NØLKE, H. (Éd.). *Macro-syntaxe et macro-sémantique*. Actes du Colloque International d'Århus [17-19 mai 2001]. Bern: Peter Lang, 2002. p. 95-118.

BÖGELS, S.; SCHRIEFERS, H.; VONK, W.; CHWILLA, D. J.; KERKHOFS, R. The Interplay Between Prosody and Syntax in Sentence Processing: The Case of Subject- and Object-Control Verbs. *Journal of Cognitive Neuroscience*, Cambridge, v. 22, n. 5, p. 1036-1053, 2010.

BÖGELS, S.; SCHRIEFERS, H.; VONK, W.; CHWILLA, D.; KERKHOFS, R. Processing Consequences of Superfluous and Missing Prosodic Breaks in Auditory Sentence Comprehension. *Neuropsychologia*, Oxford, v. 51, p. 2715-2728, 2013.

BÖGELS, S.; TORREIRA, F. Listeners Use Intonational Phrase Boundaries to Project Turn Ends in Spoken Interaction. *Journal of Phonetics*, [s.l.], v. 52, p. 46-57, 2015.

BOLINGER, D. Pitch Accent and Sentence Rhythm. In: ABE, I.; KANEKIYO, T. (Ed.). *Forms of English: Accent, Morpheme, Order*. Cambridge, Mass: Harvard University Press, 1965. p. 139-180.

BOSSAGLIA, G.; MELLO, H.; RASO, T. Insubordination and the Syntax/Prosody Interface in Spoken Brazilian Portuguese: Data on Adverbial Clauses. In: IZRE'EL, S.; MELLO, H.; PANUNZI, A.; RASO, T. *In Search for a Reference Unit of Spoken Language: A Corpus Driven Approach*. Amsterdam: John Benjamins. (Forthcoming).

BUHMANN, J.; CASPERS, J.; HEUVEN, V. J. van; HOEKSTRA, H.; MARTENS, J-P.; SWERTS, M. Annotation of Prominent Words, Prosodic Boundaries and Segmental Lengthening by Non-Expert Transcribers in the Spoken Dutch Corpus. In: LREC, 3rd, 2002, Las Palmas. *Proceedings...* Las Palmas: ELRA, 2002. p. 779-785.

BYBEE, J. *Language, Usage and Cognition*. Cambridge: CUP, 2010.

CARLSON, R.; HIRSCHBERG, J.; SWERTS, M. Cues to Upcoming Swedish Prosodic Boundaries: Subjective Judgment Studies and Acoustic Correlates. *Speech Communication*, [s.l.], v. 46, p. 326-333, 2005. Doi: <https://doi.org/10.1016/j.specom.2005.02.013>

CAVALCANTE, F. *The Topic Unit in Spontaneous American English: a Corpus-Based Study*. 2016. Master (Thesis) – Universidade Federal de Minas Gerais, Belo Horizonte, 2016.

CHAFE, W. *Discourse, Consciousness and Time. The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: Chicago University Press, 1994.

COUPER-KUHLEN, E. Prosody and sequence organization in English conversation. In: COUPER-KUHLEN, E.; FORD, C. E. (Ed.). *Sound Patterns in Interaction: Cross-Linguistic Studies from Conversation*. Amsterdam: John Benjamins, 2004. p. 335-376.

COWAN, N. *Attention and Memory: An Integrated Framework*. Oxford: Oxford University Press, 1998.

CRESTI, E. *Corpus di italiano parlato*. Firenze: Accademia della Crusca, 2000. 2 v.

CRESTI, E. Notes on Lexical Strategy, Structural Strategies and Surface Clause Indexes in the C-ORAL-ROM Spoken Corpora. In: CRESTI, E.; MONEGLIA, M. (Ed.). *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam; Philadelphia: John Benjamins, 2005. p. 209-256.

CRESTI, E. Syntactic Properties of Spontaneous Speech in the Language into Act Theory: Data on Italian Complements and Relative Clauses. In: RASO, T.; MELLO, H. (Ed.). *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins, 2014.

CRESTI, E. The Pragmatic Analysis of Speech and Its Illocutionary Classification According to Language into Act Theory. In: IZRE'EL, S.; MELLO, H.; PANUNZI, A.; RASO, T. *In Search for a Reference Unit of Spoken Language: A Corpus Driven Approach*. Amsterdam: John Benjamins. (Forthcoming).

CRESTI, E.; MONEGLIA, M. (Ed.). *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: John Benjamins, 2005.

CROFT, W. Intonational Units and Grammatical Structure. *Linguistics*, [s.l.], v. 33, n. 5, p. 839-882, 1995.

CRUTTENDEN, Alan. *Intonation*. 2nd edition. New York: Cambridge University Press, 1997.

CRYSTAL, D. *Prosodic Systems and Intonation in English*. Cambridge: CUP, 1969.

DILLEY, L.; SHATTUCK-HUFNAGEL, S.; OSTENDORF, M. Glottalization of Word-Initial Vowel as a Function of Prosodic Structure. *Journal of Phonetics*, [s.l.], v. 24, p. 423- 444, 1996.

DRURY, J. E.; BAUM, Sh. R.; VALERIOTE, H.; STEINHAUSER, K. Punctuation and Implicit Prosody in Silent Reading: An ERP Study Investigating English Garden-Path Sentences. *Frontiers in Psychology*, [s.l.], Sept. 2016. Doi: <https://doi.org/10.3389/fpsyg.2016.01375>

DU BOIS, J. W.; CHAFE, W. L.; MEYER, Ch.; THOMPSON, S. *Santa Barbara Corpus of Spoken American English*. Washington DC: Linguistic Data Consortium, 2000-2005.

DU BOIS, J. W.; CUMMING, S.; SCHUETZE-COBURN, S.; PAOLINO, D. Discourse transcription. *Santa Barbara Papers in Linguistics*, Santa Barbara, v. 4, n. 1, p. 225, 1992.

EVANS, N.; WATANABE, H. The dynamics of insubordination: An overview. In: _____. (Ed.). *Insubordination*. Amsterdam: John Benjamins, 2016.

FERY, C. *Intonation and Prosodic Structure*. Cambridge: CUP, 2017.

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.

FON, J.; JOHNSON, K.; CHEN, S. Durational Patterning at Syntactic and Discourse Boundaries in Mandarin Spontaneous Speech. *Language and Speech*, [s.l.], v. 54, n. 1, p. 5-32, 2011.

FUCHS, S.; KRIVOKAPIĆ, J.; JANNEDY, S. Prosodic Boundaries in German: Final Lengthening in Spontaneous Speech. *The Journal of the Acoustical Society of America*, [s.l.], v. 127, n. 3, p. 1851, 2010. Doi: 10.1121/1.3384378

FURHT, B.; VILLANUSTRE, F. *Big Data Technologies and Applications*. [s.l.]: Springer, 2016.

GAROFOLO, J. S.; LAMEL, L. F.; FISHER, W. M.; FISCUS, J. G.; PALLETT, D. S.; DAHLGREN, N. L.; ZUE, V. TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1. Web Download. Philadelphia: Linguistic Data Consortium, 1993.

GEE, J. P.; GROSJEAN, F. Performance Structures: A Psycholinguistic and Linguistic Appraisal. *Cognitive Psychology*, [s.l.], v. 15, n. 4, p. 411-458, 1983.

GLUSHKO, A.; STEINHAEUER, K.; De PRIEST, J.; KOELSCH, S. Neurophysiological Correlates of Musical and Prosodic Phrasing: Shared Processing Mechanisms and Effects of Musical Expertise. *PLoS ONE*, San Francisco, v. 11, n. 5, 2016.

GORDON, M.; LADEFOGED, P. Phonation Types: a Cross-Linguistic Overview. *Journal of Phonetics*, [s.l.], v. 29, p. 383-406, 2001.

GROSJEAN, F.; DOMMERGUES, J. Y. Les structures de performance en psycholinguistique. *L'Année Psychologique*, [s.l.], v. 83, n. 2, p. 513-536, 1983.

GROSJEAN, F.; GROSJEAN, L.; LANE, H. The Patterns of Silence: Performance Structures in Sentence Production. *Cognitive Psychology*, [s.l.], v. 11, n. 1, p. 58-81, 1979.

HANSON, H. M.; CHUANG, E. S. Glottal Characteristics of Male Speakers: Acoustic Correlates and Comparison with Female Data. *Journal of the Acoustical Society of America*, [s.l.], v. 106, n. 2, p. 1064-1077, 2001.

t'HART, J. Differential Sensitivity to Pitch Distance, Particularly in Speech. *The Journal of the Acoustical Society of America*, [s.l.], v. 69, n. 3, p. 811-821, 1981.

t'HART, J.; COLLIER, R.; COHEN, A. *A Perceptual Study on Intonation: An Experimental Approach to Speech Melody*. Cambridge: CUP, 1990.

HICKOK, G.; POEPEL, D. Towards a Functional Neuroanatomy of Speech Perception. *Trends in Cognitive Sciences*, [s.l.], v. 4, n. 4, p. 131-138, 2000.

HIRST, D.; Di CRISTO, A. A Survey of Intonation Systems. In: _____. (Ed.). *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press, 1998.

HUGGINS, A. W. F. Just Noticeable Differences for Segment Duration in Natural Speech. *The Journal of the Acoustical Society of America*, [s.l.], v. 51, n. 4B, p. 1270-1278, 1972.

HWANG, H.; STEINHAEUER, K. Phrase Length Matters: The Interplay Between Implicit Prosody and Syntax in Korean 'Garden Path' Sentences. *Journal of Cognitive Neuroscience*, Cambridge, v. 23, n. 11, p. 3555-3575, 2011.

HYAMS, N.; ORFITELLI, R. The Acquisition of Syntax. In: CAIRNS, H.; FERNANDEZ, E. (Ed.). *Handbook of Psycholinguistics*. [s.l.]: Wiley; Blackwell Publishers, 2015.

IZRE'EL, S. The Corpus of Spoken Israeli Hebrew: Textual Samples. *Lesson Enu*, v. 64, p. 289-314, 2002.

IZRE'EL, S.; MELLO, H.; PANUNZI, A.; RASO, T. (Ed.). *In Search for a Reference Unit of Spoken Language: A Corpus Driven Approach*. Amsterdam: John Benjamins. (Forthcoming)

KLATT, D. H.; COOPER, W. E. Perception of Segment Duration in Sentence Contexts. In: COHEN, A.; NOOTEBOOM, S. G. (Ed.). *Structure and Process in Speech Perception*. Berlin; Heidelberg: Springer, 1975. p. 69-89.

KOFFI, E. A Just Noticeable Difference (JND) Reanalysis of Fry's Original Acoustic Correlates of Stress in American English. *Linguistic Portfolios*, St. Cloud, v. 7, 2018.

LADD, D. R. *Intonational phonology*. Cambridge: Cambridge University Press, 1996.

LIEBERMAN, P. Some Acoustic Correlates of Word Stress in American English. *Journal of the Acoustical Society of America*, [s.l.], v. 32, p. 451-454, 1960.

LINELL, P. *The Written Language Bias in Linguistics: Its Nature, Origins and Transformations*. London; New York: Routledge, 2005.

MacNEILAGE, P. F. The Frame/Content Theory of Evolution of Speech Production. *Behavioral and Brain Sciences*, Cambridge, v. 21, n. 4, p. 499-511, 1998.

MARTIN, P. Les problèmes de l'intonation: recherches et applications. *Langue Française*, [s.l.], v. 19, p. 4-32, 1973.

MARTIN, P. Prosodic and Rhythmic Structures in French. *Linguistics*, [s.l.], v. 25, n. 5, p. 925-950, 1987.

MELLO, H.; RASO, T.; MITTMANN, M. VALE, H.; CÔRTEZ, P. Transcrição e segmentação prosódica do *corpus* c-oral-brasil: critérios de implementação e validação. In: RASO, T.; MELLO, H. (Ed.). C-ORAL-BRASIL I. *Corpus* de referência do português brasileiro falado informal. Belo Horizonte: UFMG, 2012. p. 125-174.

METTOUCHI, A.; CAUBET, D.; VANHOVE, M.; TOSCO, M.; COMRIE, B.; IZRE'EL, S. CORPAFROAS, A Corpus for Spoken Afroasiatic Languages: Morphosyntactic and Prosodic Analysis. In: ITALIAN MEETING OF AFRO-ASIATIC LINGUISTICS, 13th, 2010, Padova. *Proceedings...* Padova: Sargon, 2010. p.177-180.

MITTMANN, M. M. *O C-ORAL-BRASIL e o estudo da fala informal: um novo olhar sobre o Tópico no Português Brasileiro*. 2012. Dissertation (Ph.D.) – Universidade Federal de Minas Gerais, Belo Horizonte, 2012.

MITTMANN, M. M.; BARBOSA, P. A. An Automatic Speech Segmentation Tool Based on Multiple Acoustic Parameters. *CHIMERA. Romance Corpora and Linguistic Studies*, Madrid, v. 3, p.133-147, 2016.

MO, Y. Duration and Intensity as Perceptual Cues for Naïve Listeners' Prominence and Boundary Perception. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 4th., 2008, Campinas. *Proceedings...* Campinas: ISCA, 2008. p. 739-742.

MO, Y.; COLE, J.; LEE, E. K. Naïve Listeners' Prominence and Boundary Perception. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 4th., 2008, Campinas. *Proceedings...* Campinas: ISCA, 2008. p. 735-738.

MO, Y.; COLE, J. Perception of Prosodic Boundaries in Spontaneous Speech with and Without Silent Pauses. *The Journal of the Acoustical Society of America* [s.l.], v. 127, n. 3, p. 1956, 2010.

MONEGLIA, M.; CRESTI, E. L'intonazione e i criteri di trascrizione del parlato adulto e infantile. In: BORTOLINI, U.; PIZZUTO, E. (Ed.). *Il Progetto CHILDES Italia*. Pisa: Del Cerro, 1997. p. 57-90.

MONEGLIA, M.; FABBRI, M.; QUAZZA, S.; ANDREA; PANIZZA, A.; DANIELI, M.; GARRIDO, J. M.; SWERTS, M. Evaluation of Consensus on the Annotation of Terminal and Non-Terminal Prosodic Breaks in the C-ORAL-ROM corpus. In: CRESTI, E.; MONEGLIA, M. (Ed.). *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: John Benjamins, 2005. p. 257-276.

MONEGLIA, M.; RASO, T. Notes Language into Act Theory (L-AcT). In: RASO, T.; MELLO, H. (Ed.). *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins, 2014. p. 468-495.

MONNIN, P.; GROSJEAN, F. Les structures de performance en français: caractérisation et prédiction. *L'Année Psychologique*, [s.l.], v. 93, p. 9-30, 1993.

MORAES, J. A.; RILLIARD, A. Illocution, Attitude and Prosody: A Multimodal Analysis. In: RASO, T.; MELLO, H. (Ed.). *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins, 2014. p. 233-270.

NESPOR, Marina; VOGEL, Irene. *Prosodic Phonology*. Dordrecht: Foris Publications, 1986.

NI, C. J.; ZHANG, A. Y.; LIU, W. J.; XU, B. Automatic Prosodic Break Detection and Feature Analysis. *Journal of Computer Science and Technology*, [s.l.], v. 27, n. 6, p. 1184-1196, 2012.

NICKELS, S.; OPITZ, B.; STEINHAEUER, K. ERPs Show that Classroom-Instructed Late Second Language Learners Rely on the Same Prosodic Cues in Syntactic Parsing as Native Speakers. *Neuroscience Letters*, [s.l.], v. 557, p. 107-111, 2013.

OSTENDORF, Mari; PRICE, Patti; SHATTUCK-HUFNAGEL, Stefanie. Boston University Radio Speech Corpus LDC96S36. DVD. Philadelphia: Linguistic Data Consortium, 1996.

PAUKER, E.; ITZHAK, I.; BAUM, S. R.; STEINHAEUER, K. Effects of Cooperating and Conflicting Prosody in Spoken English Garden Path Sentences: ERP Evidence for the Boundary Deletion Hypothesis. *Journal of Cognitive Neuroscience*, Cambridge, v. 23, n. 10, p. 2731-2751, 2011.

PIERREHUMBERT, J. B. *The Phonology and Phonetics of English Intonation*. 1980. Dissertation (Doctoral) – Massachusetts Institute of Technology, 1980.

PIETRANDREA, P.; KAHANE, S.; LACHERET, A.; SABIO, F. In: RASO, T.; MELLO, H. (Ed.). *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins, 2014.

PIKE, K. *The Intonation of American English*. Ann Arbor: University of Michigan Press, 1945.

PRICE, P. J.; OSTENDORF, M.; SHATTUCK-Hufnagel, S.; FONG, C. The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America*, [s.l.], v. 90, n. 6, p. 2956-2970, 1991.

RASO, T.; CAVALCANTE, F.; MITTMANN, M. M. Prosodic Forms of the Topic Information Unit in a Cross-Linguistic Perspective: a First Survey. In: GSCP INTERNATIONAL CONFERENCE, 2016, Napole. *Proceedings...* Napole. (Forthcoming).

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

RASO, T.; MELLO, H. (Ed.). *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins, 2014.

RASO, T.; MELLO, H. (Ed.). The C-ORAL-BRASIL II. *Corpus de referência do português falado (formal em contexto natural, mídia e telefone)*. (Forthcoming).

RASO, T.; MITTMANN, M. As principais medidas da fala. In: RASO, T.; MELLO, H. (Ed.). C-ORAL-BRASIL I. *Corpus de referência do português brasileiro falado informal*. Belo Horizonte: UFMG, 2012. p. 177-220.

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*, Madrid, v. 3, p. 221-249, 2016.

REDI, L.; SHATTUCK-HUFNAGEL, S. Variation in the Realization of Glottalization in Normal Speakers. *Journal of Phonetics*, [s.l.], v. 29, p. 407-29, 2001.

RIETVELD, A.; GUSSENHOVEN, C. On the Relation Between Pitch Excursion Size and Prominence. *Journal of Phonetics*, [s.l.], v. 13, p. 299-308, 1985.

ROBIN, D. A. *et al.* Auditory Perception of Temporal and Spectral Events in Patients with Focal Left and Right Cerebral Lesions. *Brain and Language*, [s.l.], v. 39, p. 539-555, 1990.

ROCHA, B.; RASO, T. O pronome lembrete e a Teoria da Língua em Ato: uma análise baseada em corpora. *Veredas*, Juiz de Fora, v. 17, n. 2, p. 39-59, 2013.

SCHEGLOFF, E. A. Reflections on Studying Prosody in Talk-in-Interaction. *Language and Speech*, [s.l.], v. 41, n. 3-4, p. 235-263, 1998.

SELKIRK, E. Sentence prosody: Intonation, Stress and Phrasing. In: GOLDSMITH, J. A. (Ed.). *The Handbook of Phonological Theory*. Oxford: Blackwell, 1995. p. 550-569.

SHAH, A. P.; BAUM, S. R.; DWIVEDI, V. D. Neural Substrates of Linguistic Prosody: Evidence from Syntactic Disambiguation in the Productions of Brain-Damaged Patients. *Brain and Language*, [s.l.], v. 96, p. 78-89, 2006.

SHRIBERG, E.; STOLCKE, A.; HAKKANI-TÜR, D.; TÜR, G. Prosody-Based Automatic Segmentation of Speech into Sentences and Topics. *Speech Communication*, [s.l.], v. 32, n. 1-2, p. 127-154, 2000.

SCHUURMAN, I.; SCHOUPPE, M.; HOEKSTRA, H.; Van der Wouden, T. CGN, an Annotated Corpus of Spoken Dutch. In: INTERNATIONAL WORKSHOP ON LINGUISTICALLY INTERPRETED CORPORA (LINC-03), 4TH, 2003, Budapest. *Proceedings...* Budapest: EAACL, 2003.

SIGNORINI, S. *Topic e soggetto in corpora di italiano parlato spontaneo*. 2004. Dissertation (Ph.D.) – Università Firenze, Firenze, 2004.

STEINHAEUER, K.; ALTER, K.; FRIEDERICI, A. D. Brain Potentials Indicate Immediate Use of Prosodic Cues in Natural Speech Processing. *Nature Neuroscience*, [s.l.], v. 2, n. 2, p. 191-196, 1999.

STEINHAEUER, K. Electrophysiological Correlates of Prosody and Punctuation. *Brain and Language*, [s.l.], v. 86, n. 1, Special issue on the Neuronal Basis of Language, p. 142-164, 2003.

STEINHAUER, K.; FRIEDERICI, A. D. Prosodic Boundaries, Comma Rules, and Brain Responses: The Closure Positive Shift in ERPs as a Universal Marker for Prosodic Phrasing in Listeners and Readers. *Journal of Psycholinguistic Research*, [s.l.], v. 30, n. 3, p. 267-295, 2001.

SWERTS, M. Prosodic Features of Discourse Units. 1994. Thesis (PhD) – Technische Universiteit Eindhoven, 1994.

SWERTS, M. Prosodic Features at Discourse Boundaries of Different Strength. *The Journal of the Acoustical Society of America*, [s.l.], v. 101, n. 1, p. 514-521, 1997.

SWERTS, M.; COLLIER, R.; TERKEN, J. Prosodic Predictors of Discourse Finality in Spontaneous Monologues. *Speech Communication*, [s.l.], v. 15, n. 1-2, p. 79-90, 1994.

TEIXEIRA FALCÃO, B. H. *Correlatos fonético-acústicos de fronteiras prosódicas na fala espontânea*, 2017. Thesis (Master) – Universidade Federal de Minas Gerais, Belo Horizonte, 2017.

TESNIÈRE, L. *Éléments de Syntaxe Structurale*. Paris: C. Klincksieck, 1965.

THORNTON, R. Children's Acquisition of Syntactic Knowledge. In: ARONOFF, M. (Ed.). *Oxford Research Encyclopedia of Linguistics*, 2016.

TSENG, C. Y.; CHANG, C. H. Pause or No Pause? Prosodic Phrase Boundaries Revisited. *Tsinghua Science and Technology*, [s.l.], v. 13, n. 4, p. 500-509, 2008.

TSENG, C. Y.; FU, B. Duration, Intensity and Pause Predictions in Relation to Prosody Organization. In: EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY, INTERSPEECH, 9th, 2005. Lisboa. *Proceedings...* Lisboa: SISCA, 2005. p. 1405-1408. Available at: <<http://www.ling.sinica.edu.tw/eip/FILES/publish/2007.4.12.99500673.0143164.pdf>>. Retrieved on: Dec. 1, 2016.

TYLER, J. Prosodic Correlates of Discourse Boundaries and Hierarchy in Discourse Production. *Lingua*, [s.l.], v. 133, p. 101-126, 2013.

Van DONZEL, M. E. *Prosodic Aspects of Information Structure in Discourse*. Den Haag: Holland Academic Graphics/IFOTT, 1999.

WAGNER, A. Acoustic Cues for Automatic Determination of Phrasing. In: SPEECH PROSODY 2010 – INTERNATIONAL CONFERENCE, 5th., 2010, Chicago. *Proceedings...* Chicago: ISCA, 2010. Available at: <https://www.isca-speech.org/archive/sp2010/papers/sp10_196.pdf>. Retrieved on: Sept. 2018.

WAGNER, M.; WATSON, D. G. Experimental and Theoretical Advances in Prosody: A Review. *Language and Cognitive Processes*, [s.l.], v. 25, n. 7-9, p. 905-945, 2010.

WIGHTMAN, C. W.; SHATTUCK-HUFNAGEL, S.; OSTENDORF, M.; PRICE, P. J. Segmental Durations in the Vicinity of Prosodic Phrase Boundaries. *The Journal of the Acoustical Society of America* [s.l.], v. 91, n. 3, p. 1707–1717, 1992.

XU, Y. In Defense of Lab Speech. *Journal of Phonetics*, [s.l.], v. 38, n. 3, p. 329-336, 2010.

YANG, Y.; WANG, B. Acoustic Correlates of Hierarchical Prosodic Boundary in Mandarin. In: SPEECH PROSODY, 2002, Aix-en-Provence. *Proceedings...* Aix-en-Provence: Laboratoire Parole et Langage, 2002.

YOON, T-J.; CHAVARRÍA, S.; COLE, J.; HASEGAWA-JOHNSON, M. Intertranscriber Reliability of Prosodic Labeling on Telephone Conversation Using ToBI. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 2004. Nara, Japan. *Proceedings...* Nara: ISCA, 2004. p. 2722-2732.

ZATORRE, R. J. Cerebral Correlates of Human Auditory Processing: Perception of Speech and Musical Sounds. In: SYKA, J. (Ed.). *Acoustical Signal Processing in the Central Auditory System*. Prague: Plenum Press, 1997. p. 453-468.



A segmentação da fala espontânea: aspectos prosódicos, funcionais e aplicações para a tecnologia

Spontaneous Speech Segmentation: Functional and Prosodic Aspects with Applications for Automatic Segmentation

Plínio A. Barbosa

Universidade Estadual de Campinas, Campinas, São Paulo / Brazil

pabarbosa.unicampbr@gmail.com

Tommaso Raso

Universidade Federal de Minas Gerais, Belo Horizonte. Minas Gerais /Brazil

tommaso.raso@gmail.com

Este número da *Revista de Estudos da Linguagem* é dedicado a um tema enfrentado em diversas outras iniciativas promovidas pelos organizadores junto com colegas de outros países. O tema, que nos últimos anos tem adquirido um papel cada vez mais importante nas disciplinas que estudam a produção e a percepção da fala, é a segmentação da fala em unidades menores, vista sob perspectivas tanto formais quanto funcionais, tanto fundamentalmente teóricas quanto com foco mais empírico. Entre as principais iniciativas, citamos:

- Dois workshops internacionais (o IV Leel e o X Lablita International workshop *Unit of Reference for Spontaneous Speech and their Correlation Across Language*, realizado em agosto de 2015 na UFMG; e o workshop *Spoken Corpora advances: prosody as the crux of speech segmentation, annotation and multilevel linguistic studies*, organizado na Cidade do Cabo em junho de 2018, dentro do 20º International Congress of Linguists);

- O livro *In Search for a Reference Unit of Spoken Language: A Corpus Driven Approach*, a ser lançado em breve pela editora John Benjamins;
- Um número especial do *Journal of Speech Sciences* programado para sair em meados de 2019.

Todas essas iniciativas são dedicadas ao tema da segmentação prosódica da fala, um tema que tem cada vez mais se tornado central para entender a estruturação da fala em diversos níveis, bem como a relação dessa estruturação com as funções comunicativas da linguagem. As disciplinas interessadas no tema, e a linguística *in primis*, têm se desenvolvido enormemente tanto a partir da contribuição dos avanços tecnológicos e da estatística aplicados aos estudos linguísticos, quanto da contribuição dos avanços das próprias teorias linguísticas. De fato, até recentemente, o estudo da segmentação da fala considerava a segmentação, quase exclusivamente, da chamada fala de laboratório (*lab speech*), que inclui a fala lida e a fala eliciada sob várias formas (XU, 2010), a partir da manipulação de eventos externos pelo pesquisador (como pela proposta de tarefas com um ou mais participantes como *map task* e jogos eletrônicos, pela condução de entrevistas sobre temas específicos, *inter alia*). Há alguns anos, no entanto, tornou-se possível abordar a fala sem roteiro prévio (*non scripted speech*) extraída de *corpora* de fala espontânea em situações comunicativas naturais variadas e com boa qualidade acústica. Nesse artigo introdutório ao número temático apresentamos um panorama, mesmo que parcial, das questões científicas em jogo, dos resultados alcançados até aqui e dos passos que já se anunciam para o futuro.

1. Segmentação prosódica: entre forma e função

Contrariamente à escrita, que é um produto que permanece no tempo e no espaço, a fala é um processo cujo resultado desaparece logo após a sua manifestação, se não consideramos nesse exame as tecnologias que realizam seu registro. Apenas permanecem como subproduto imediato algumas consequências cognitivas do discurso, mas não a fala em si (LINELL, 2005; BLANCHE-BENVENISTE; JEANJEAN, 1987). Ausente na escrita em sua manifestação acústica, a não ser por meros indicativos pelos sinais de pontuação, a prosódia é o componente essencial para os estudos de segmentação da fala. Hoje é possível, graças

à tecnologia e a softwares dedicados, reproduzir a fala por quantas vezes se achar necessário e realizar procedimentos de anotação que permitem delimitar diferentes unidades para poder estudá-la: sílabas, grupos de sílabas ou palavras, unidades prosódicas de diferente dimensão e estatuto teórico, bem como sequências de enunciados de interesse. Isso permite a observação sistemática e a medição de muitos aspectos da fala que, sem a tecnologia, haviam sido em certa medida apenas intuídos através da sensibilidade auditiva dos precursores da prosódia contemporânea (cf. PIKE, 1945; LIEBERMAN, 1960; BOLINGER, 1965), mas que não podiam ser nem aprofundados nem demonstrados. Entre esses aspectos, um lugar de importância primária é ocupado pelos diferentes componentes em que é possível segmentar o fluxo da fala e pela reflexão sobre suas formas e funções. Enfim, se tornou possível tentar a reconstrução da complexa estrutura prosódica (e não somente) da fala humana.

Além disso, a tecnologia tornou possível compilar e investigar grandes quantidades de dados de fala, tratados e anotados de diferentes maneiras e adequados especificamente a estudos dos mais variados, numa linha de pensamento que privilegia a obtenção de conhecimento a partir de *corpora* bastante extensos, no que se convencionou chamar hoje de “*big data*” (cf. FURHT; VILLANUSTRE, 2016). O tratamento informatizado do sinal acústico nos permite segmentar o discurso em unidades menores, desde o enunciado (ou talvez desde unidades maiores como os “parágrafos”) até a sílaba e os seus constituintes; e nos permite investigar como a fala humana veicula as fronteiras (ou a ausência delas) em diferentes níveis hierárquicos.

Dependendo do interesse de estudo, a fala pode ser segmentada em unidades de diferentes tamanhos e naturezas, cada uma delas sendo capaz de mostrar algumas de suas propriedades e cada uma delas sendo delimitada por algum tipo de fronteira. Considerando apenas as unidades acima do nível da palavra, podemos dividir a fala em grupos acentuais (ou pés *n*-ários, os grupos de sílabas até uma tônica, no caso de línguas com cabeça à direita), em unidades prosódicas chamadas de entonacionais ou tonais ou de grupos prosódicos, em enunciados, ou, em uma perspectiva de natureza sintática, em sintagmas entoacionais (IP), orações e sentenças. Cada tipo de segmentação está direta ou indiretamente associado a uma visão teórica, mas em muitos casos isso não impede uma investigação empírica, cujos resultados podem ser analisados à luz de diferentes perspectivas teóricas. Nos últimos anos, vários *corpora* com a anotação

prosódica da fronteira foram compilados para diferentes línguas (AURAN *et al.*, 2004; DU BOIS *et al.*, 2000-2005; OSTENDORF *et al.*, 1996; CRESTI; MONEGLIA, 2005; SCHUURMAN *et al.*, 2003; IZRE'EL, 2002; RASO; MELLO, 2012 e em preparação; METTOUCHI *et al.*, 2010; GAROFOLO, *et al.*, 1993,).

Qualquer segmentação implica a presença de uma fronteira, seja ela de fato percebida ou proposta teoricamente. Sendo assim, a fronteira pode ser entendida como uma ruptura fisicamente percebida; pode se referir a um limite verificável para a realização de fenômenos linguísticos; e pode ainda ser considerada em uma região entre duas unidades, sendo que essa região pode ser de natureza perceptível ou não.

Este número temático busca estudar a segmentação enfocando o que pode ser considerado como a unidade de referência do processo da fala (IZRE'EL *et al.*, no prelo). A própria noção de unidade de referência pode ser entendida de diferentes maneiras, mas provisoriamente podemos defini-la como uma unidade mínima de sentido completo e autônomo comunicativamente que compõe um texto falado (CRESTI, 2000; MONEGLIA; RASO, 2014). Essa definição não é incontroversa, mas nos permite começar a busca.

Todos os tipos de unidades nomeadas acima, independentemente de como são definidos, são separados por fronteiras que são definidas dando um peso maior ou menor à percepção ou à teoria; dificilmente um dos dois critérios de individualização exclui completamente o outro. Nos trabalhos deste número temático está sempre presente uma base perceptiva, mas algumas contribuições dão um peso maior que outras aos aspectos teóricos, sendo que esses aspectos variam de uma contribuição para outra. Com essas diferenças de perspectiva muda também o conceito de fronteira.

De natureza teórica são as fronteiras de constituintes de abordagem sintática ou informacional; isso não significa que elas não possam ser associadas a fronteiras de caráter prosódico, que constituem o interesse primário desse número temático. De fato, entendemos que a prosódia guia a interpretação sintática, como em casos como o da sentença “A ovelha de raça brasileira.” A partir dessa unidade da escrita, dois enunciados podem ser emitidos segundo duas formas de agrupamento distintas, em que “/” representa uma fronteira não terminal forte:

[A ovelha de raça] / [brasileira] vs. [A ovelha] / [de raça brasileira]

No primeiro caso se trata de uma ovelha de raça não informada nascida no Brasil e, no segundo caso, de uma ovelha que é de raça desenvolvida no Brasil. São assim os constituintes prosódicos que permitem a escansão adequada da estrutura sintática de cada sentença. Isto é, a prosódia permite a desambiguação entre as duas interpretações possíveis, pois os limitados recursos da escrita não permitem resolver a distinção. Nesse caso, a prosódia guia a interpretação sintática e os constituintes sintáticos e prosódicos são consequentemente congruentes, isto é, têm os mesmos limites. Por conta disso os autores deste número temático que tratam diretamente da questão da segmentação da fala consideram unidades que são constituintes prosódicos.

Assim, quase todas as contribuições aqui estudam a organização da fala em unidades que podem ser consideradas como extensíveis a unidades entonacionais. Por conta disso, nesta apresentação utilizaremos a expressão “unidade entonacional” de forma geral, mas deixamos claro que essa unidade se estrutura com base em parâmetros que não incluem apenas a frequência fundamental (f_0), mas também parâmetros de natureza duracional, de intensidade e possivelmente de qualidade de voz. Um único trabalho (o de Ph. Martin) segmenta a fala em grupos acentuais, o que não exclui o fato de um grupo acentual ou um conjunto de grupos acentuais coincidir com a unidade entonacional. A segmentação em grupos acentuais pode, portanto, ser vista também como a oportunidade de investigar a estrutura interna da unidade entonacional, enriquecendo assim, e não contradizendo, as perspectivas que preferem se concentrar na análise da unidade entonacional.

É difícil definir a unidade entonacional sem fazer referência ou à percepção ou a um postulado de natureza teórica. Mas em geral a unidade entonacional é definida como o grupo de palavras (pode ser também uma única palavra e, em casos mais raros em que entra em jogo a ênfase em sílabas, menos de uma palavra. Nesse último caso, a fronteira é uma consequência perceptiva da proeminência da unidade) delimitado entre uma fronteira prosódica e outra, gerando um contorno entonacional coerente e separado fisicamente e perceptivamente dos contornos precedente e seguinte (DU BOIS *et al.*, 1992, p. 17; CRUTTENDEN, 1997). Essa definição mascara algumas dificuldades em capturar as propriedades de uma unidade entonacional sem fazer referência à fronteira, e por sua vez sem identificar a fronteira a partir do conceito de unidade entonacional, com evidente risco de circularidade. A própria

definição de “contorno coerente” não é completamente satisfatória, já que não sabemos com clareza quais são os parâmetros que permitem ou rompem a coerência.

Do ponto de vista funcional, a unidade entonacional pode ser estudada e definida linguisticamente com base em perspectivas diferentes. As principais são a perspectiva sintática, a perspectiva informacional (CHAFE, 1994; RASO; MELLO, 2014) e a perspectiva conversacional (BARTH-WEINGARTEN, 2016). Mas a própria individualização da unidade entonacional é problemática. De fato, não é sempre óbvio reconhecer um perfil prosódico coerente ou uma fronteira prosódica. No que tange a identificação de uma fronteira, geralmente os estudos se baseiam no acordo estatístico entre segmentadores: um determinado trecho de fala é oferecido para um certo número de segmentadores e se compara o acordo que eles tiveram em segmentar, de oitiva, o trecho em unidades menores. Outras abordagens consideram a percepção associada a alguns traços formais visíveis a partir de ferramenta de software, como o que é chamado de tom de fronteira (*boundary tone*), um determinado movimento de f0 alinhado ao final da unidade, na linha de investigação da Fonologia Métrica-Autossegmental (LADD, 1996; PIERREHUMBERT, 1980).

Testes estatísticos de análise da coerência entre avaliadores (segmentadores) mostram que o acordo na identificação das fronteiras, e por consequência das unidades, é muito alto (superior a 80 %, especialmente para o caso das fronteiras terminais; MELLO *et al.*, 2012; MONEGLIA *et al.*, 2005; YOON *et al.*, 2004; BUHMANN *et al.*, 2002). É, portanto, consensual que um importante nível de organização da fala seja constituído pela unidade entonacional. As razões dessa organização, ao contrário, são controversas: segundo alguns autores (cf. COWAN, 1998) essa segmentação do fluxo da fala é devida aos limites de memória, que impõem agrupamentos de um número limitado de sílabas para o processamento linguístico. Segundo outros, as unidades seriam devidas a motivações cognitivas (CHAFE, 1994; CROFT, 1995; BYBEE, 2010). Segundo outros ainda, a segmentação corresponde a unidades de natureza sintática e, portanto, fronteira prosódica e fronteira sintática seriam correlacionadas, especialmente nas abordagens de natureza fonológica da prosódia que pressupõem um mapeamento entre constituintes sintáticos e os limites de unidades prosódicas (NESPOR; VOGEL, 1986; SELKIRK, 1995). Uma quarta proposta, dominante nesse número temático, atribui à fronteira prosódica o valor de delimitar unidades de natureza

informativa, independentemente de sua organização sintática. Outros ainda vêm uma correspondência entre a prosódia e unidades de outro domínio discursivo (COUPER-KUHLEN, 2004; SCHEGLOFF, 1998). Quem estuda a prosódia como correlacionada a domínios linguísticos de natureza não sintática tende também a considerar a prosódia como um elemento estrutural implementado antes dos elementos segmentais (cf. a teoria *Frame/Content* de MacNEILAGE, 1998). Uma visão interessante dentro dos estudos prosódicos (HIRST; Di CRISTO, 1998; BARBOSA, 2006) tenta uma conciliação entre constituintes sintáticos e prosódicos, propondo que a sintaxe se limitaria a impor algumas restrições, mas não determinaria a posição das fronteiras: estas poderiam, portanto, aparecer somente em posições compatíveis com a estruturação sintática, sem por isso marcar necessariamente constituintes dessa natureza, já que, dada uma mesma sentença, seriam várias as posições compatíveis com a estruturação sintática onde poderia ser colocada uma fronteira, com cada posição sinalizando uma interpretação cognitivo-informativa diferente. Por outro lado, muitos estudiosos da sintaxe estão percebendo como a prosódia é essencial para o funcionamento de estruturas que apresentam fortes dificuldades para as explicações sintáticas tradicionais. É o caso do fenômeno da assim chamada de insubordinação (EVANS; WATANABE, 2016; BOSSAGLIA *et al.*, no prelo). Nesses casos, a interpretabilidade da estrutura depende em maneira decisiva da sua codificação prosódica.

2. As principais questões metodológicas

As pesquisas realizadas mostram também que o estudo das fronteiras prosódicas depende da tipologia de fala e em parte da tipologia do texto falado. De fato, até recentemente, as pesquisas se concentraram no estudo da segmentação prosódica em textos lidos ou sequências limitadas executadas em laboratório, com resultados interessantes, mas que não parecem ser comparáveis com o que acontece na fala espontânea, objetivo prioritário desse número temático. É frequente em estudos de prosódia ligados à sintaxe e à fonologia que a fala de laboratório seja usada para testar relações entre prosódia e sintaxe (como no caso de desambiguação ou na investigação dos possíveis tipos de constituintes isolados por fronteiras), apresentando um número muito menor de variáveis do que a fala espontânea e uma maior previsibilidade (PRICE *et al.*, 1991). Ressalta-se ainda que, quando essa fala é lida, ela é a

realização sonora de um texto escrito, e, portanto, estruturado com base em princípios diferentes daqueles da fala espontânea.

Recentemente, alguns trabalhos sobre a fala espontânea obtiveram bons resultados na investigação dos mecanismos de segmentação, seja observando um acordo alto (maior que 80%) entre os segmentadores humanos nessa tarefa (MELLO *et al.*, 2012; MONEGLIA *et al.*, 2005; TEIXEIRA FALCÃO, 2017), seja criando softwares capazes de segmentar automaticamente textos alcançando resultados altamente comparáveis com as tarefas realizadas pelos humanos (AVANZI *et al.*, 2008; NI *et al.*, 2012; BARBOSA, 2016).

A criação de softwares capazes de automatizar a segmentação prosódica em unidades entonacionais (cf. MITTMAN; BARBOSA, 2016) só é possível porque a investigação dos parâmetros acústicos responsáveis pela percepção de fronteira tem avançado muito, graças também aos trabalhos realizados sobre a fala lida e sobre sequências realizadas em laboratório que permitiram uma primeira compreensão dos fenômenos em jogo, que são altamente complexos. Parece de fato que os parâmetros responsáveis pela nossa percepção de fronteira são diversos, que nem sempre são todos co-presentes, que seu peso pode variar dependendo das línguas e das circunstâncias da fala, levando também a questionar se é possível falar de fronteira como categoria homogênea ou se não é o caso de falar de tipos diferentes de fronteiras.

Na literatura os parâmetros que são mais mencionados são tanto de frequência fundamental (f_0), quanto de duração e de intensidade, além de parâmetros que se referem à qualidade de voz (BARTHWEINGARTEN, 2016; MO *et al.*, 2008; WAGNER; WATSON, 2010), especialmente laringalização, *creaky voice* em inglês (DILLEY *et al.*, 1996; GORDON; LADEFOGED, 2001; REDI; SHATTUCK-HUFNAGEL, 2001; HANSON *et al.*, 2001; CARLSON *et al.*, 2005). Os principais são os seguintes: a pausa silenciosa, que aqui simplesmente chamaremos de “pausa” (adiante falaremos da pausa preenchida), cuja presença parece automaticamente veicular a percepção de fronteira (MARTIN, 1973; SWERTS, 1997; SHRIBERG *et al.*, 2000; TSENG; CHANG, 2008; MO; COLE, 2010; TYLER, 2013); o alongamento das sílabas finais da unidade, ou seja uma redução da taxa de realização das últimas sílabas antes de uma fronteira (WIGHTMAN *et al.*, 1992; BARBOSA, 2008; MO *et al.*, 2008; FUCHS *et al.*, 2010; FON *et al.*, 2011; TYLER, 2013); a redução duracional das primeiras sílabas da

unidade, ou seja, a aceleração imediatamente após uma fronteira (AMIR *et al.* 2004; TYLER, 2013), correlacionada com fenômenos de *anacrusis*; o *reset* da curva de f_0 ; a mudança brusca de direção da curva de f_0 ; a mudança de intensidade no início da unidade prosódica (SWERTS *et al.*, 1994; TSENG; FU, 2005; MO, 2008); a laringalização (*creaky voice*) e talvez outras qualidades de voz não modais. A esses parâmetros, pelo menos para algumas línguas, devem ser acrescentados alguns fenômenos de natureza segmental. Por exemplo, para o inglês pode ser importante a soltura da oclusiva final ou a laringalização e golpe de glote de alguns segmentos finais como marcadores de fronteira.

Cada uma dessas pistas traz problemas para o pesquisador. Por exemplo, a pausa, que intuitivamente parece uma noção óbvia, não é identificada consensualmente: qual é o tempo mínimo de silêncio para considerar a presença de uma pausa? Como a presença dessa pausa afeta os outros parâmetros responsáveis pela percepção de fronteira? A pausa é um indicador também do tipo de fronteira ou não? Quanto à curva de f_0 , qual a contribuição relativa da diferença de nível de f_0 , de sua excursão, da direção de seu movimento e da taxa de sua variação? Já quando se considera a duração silábica, qual a extensão da região afetada pela fronteira, medida em número de sílabas? E se a mudança de duração envolve mais do que a simples sílaba fronteira, a mudança acontece na mesma proporção em cada sílaba envolvida ou não? Além disso, os trabalhos experimentais mostraram que, para avaliar de maneira confiável as medidas de natureza duracional, é necessária alguma forma de normalização que coloque à parte as propriedades intrínsecas dos segmentos, que, nesse caso, influem de maneira decisiva sobre a duração (BARBOSA, 2012). É preciso salientar ainda que a medida da duração propícia à análise prosódica deve considerar a realização e a fronteira das sílabas, o que envolve as noções de sílaba fonológica e sílaba fonética. A primeira é importante para a percepção da taxa de elocução, porque perpassa pela apreensão da sílaba pelo sistema cognitivo, enquanto a segunda fundamenta a produção da cadeia de fala e a organização de consoantes e vogais na sílaba produzida.

A pesquisa sobre os parâmetros acústicos que, no seu conjunto, veiculam a percepção de quebra (fronteira) deve considerar o peso ou contribuição relativa de cada pista acústica. Para isso, é importante considerar não somente que cada pista é perceptível somente se passar de um certo limiar, mas que esse limiar varia variando as outras pistas

(t'HART *et al.*, 1990). Isso significa, em primeiro lugar, que nós não somos capazes de perceber qualquer mudança de f_0 ou qualquer mudança de duração ou intensidade, mas somente as mudanças que ultrapassam um determinado limiar. Embora para cada parâmetro ou pista em isolado possamos conhecer a *Just Noticeable Difference* (JND), ou seja a variação mínima desse parâmetro que podemos perceber (cf. HUGGINS, 1972; KLATT; COOPER, 1975 para duração segmental; t'HART, 1981; RIETVELD; GUSSENHOVEN, 1985 para f_0 ; KOFFI, 2018 para intensidade), bem como a forma pela qual a JND varia com a partir da modificação de um outro parâmetro (por exemplo como percebemos a variação de intensidade em frequências diferentes), pouco sabemos ainda sobre como essas combinações complexas de parâmetros variam com relação à capacidade de veicular percepção de fronteira. Não é simples modelar o efeito de fronteira nas combinações de tantos parâmetros no fluxo da fala. De fato, não seria surpreendente se o peso de uma pista mudasse mudando as combinações de pistas nas quais está inserido, ou mudando os contextos de fala em que aparece: leitura ou fala espontânea, ou diferentes estilos de fala espontânea, ou ainda diferentes funções linguísticas das unidades delimitadas pelas fronteiras marcadas, sem considerar variações ligadas às características dos falantes.

De fato, os diversos estudos em diferentes línguas confirmam a importância das pistas apontadas acima para a percepção de fronteira, enquanto também revelam que cada uma dessas pistas atua com peso diferente para assinalar essa mesma fronteira (TEIXEIRA FALCÃO, 2017). Essa diferente hierarquia de pistas acústicas parece estar ligada às funções que um determinado parâmetro possui na língua. Por exemplo, em línguas tonais, a f_0 tem o papel de veicular funções linguísticas que em línguas não tonais são veiculadas por outros parâmetros. Nessas línguas, diferenças de f_0 realizam diferenças entre tons que servem para contrastar itens lexicais. Por também ter essa função, o peso da f_0 é afetado quando esse parâmetro é usado para marcar fronteira, sendo parâmetros duracionais e reset de f_0 mais relevantes para assinalar fronteiras (YANG; WANG, 2002). É provável que isso aconteça também com os outros parâmetros, que se comportariam de maneira diferente para assinalar fronteira prosódica a depender da importância que têm para sinalizar outras funções em uma determinada língua. Muito pouco sabemos também como varia o peso de um parâmetro dentro de uma combinação mais ampla ao marcar fronteiras de unidades funcionalmente diferentes.

Enquanto alguns estudos se concentram em investigar a oposição entre presença vs. ausência de fronteira (MO *et al.*; 2008; BARBOSA, 2010), outros investigam uma potencial diversidade entre as fronteiras. Nesse último caso, alguns autores propõem a existência de um número determinado de fronteiras, enquanto outros propõem um continuum entre presença e ausência de fronteira. Nesse segundo caso corre-se o risco de encontrar sempre algum grau de fronteira, por mínimo que seja, e de perder a oposição de fronteira vs. não fronteira, tornando extremamente difícil, se não impossível, qualquer consideração de natureza funcional.

Quem, por outro lado, considera que as fronteiras são um fenômeno gradiente, mas categórico, propõe uma gradação de força das diferentes fronteiras, que contudo são em número limitado; entre esses autores existe uma discordância sobre a quantidade de fronteiras de força diferente que é possível reconhecer e perceber (cf. BARBOSA, 2006, para uma discussão). Alguns trabalhos distinguem simplesmente entre fronteiras fortes e fracas, enquanto outros consideram possível individualizar mais de dois graus de força (cf. WIGHTMAN *et al.*, 1992, para o inglês; BARBOSA, 2006, para o português brasileiro; BARBOSA, 1994, para o francês), alguns chegando até a sete, o que se alinha com as já citadas teorias fonológicas da prosódia como de Nespor e Vogel (1986) e Selkirk (1995).

Outra possibilidade de inferir graus de força é pelo uso de máximos locais dos parâmetros acústicos que veiculam a fronteira prosódica como índices da força dessa fronteira (TEIXEIRA FALCÃO, 2017). Mesmo que os valores de máximos locais variem continuamente, é possível o uso de técnicas estatísticas de agrupamento (*clusterization techniques*) para inferir um número limitado de forças de fronteira que não é superior a quatro (cf. BARBOSA, 2006, para o PB e BARBOSA, 1994, para o francês). No trabalho para o PB, Barbosa (2006) utilizou máximos de duração de unidades de tamanho silábico normalizada por *z-score* para a obtenção de 3 a 4 níveis distintos, parcialmente correlacionados com fronteiras sintáticas obtidas pela projeção de uma árvore de dependência nos modelos da de Tesnière (1965). Os diferentes graus de força permitem estabelecer uma hierarquia de constituintes prosódicos que abrem a possibilidade da inferência da estrutura prosódica de um enunciado. Esse procedimento já fora proposto por Grosjean e colegas (GROSJEAN; GROSJEAN; LANE, 1979; GROSJEAN; DOMMERGUES, 1983; GEE; GROSJEAN, 1983) a partir da leitura em

taxas cada vez mais lentas e da análise de durações de vogais acrescidas das eventuais pausas silenciosas à sua direita e de índices de segmentação dos enunciados obtidos por testes com ouvintes. Esse procedimento pôde revelar assim a “*structure de performance*” (estrutura de performance), uma estrutura prosódica com as seguintes propriedades: constituintes de tamanho semelhante, organização hierárquica e estrutura simétrica (GROSJEAN; DOMMERGUES, 1983). Essas propriedades emergiram de duas restrições concorrentes: a tendência do locutor em respeitar a estrutura linguística da sentença e a tendência a equilibrar a extensão dos constituintes que produz (MONNIN; GROSJEAN, 1993, p. 28; MARTIN, 1987). A tendência de equilíbrio da extensão de constituintes prosódicos explicaria porque os sujeitos não agrupam sistematicamente o verbo com o sintagma nominal objeto ao pronunciarem frases do inglês, como seria previsto pela sintaxe, mas preferem agrupamentos do tipo (SV)O (GROSJEAN; GROSJEAN; LANE, 1979, p. 59).

A discussão sobre os tipos de fronteira, no entanto, não é somente de cunho quantitativo. Muitos autores distinguem entre fronteiras que veiculam percepção de conclusão prosódica e linguística (com interpretações diferentes sobre a natureza da unidade linguística concluída) e fronteiras que veiculam a percepção de continuidade discursiva, sinalizando que o segmento de discurso em curso não pode se considerar concluído, apesar de a fronteira marcar a conclusão de um constituinte, esse também de diferente natureza dependendo da abordagem teórica (MONEGLIA; CRESTI 1997; CRYSTAL, 1969; SWERTS, 1994; SWERTS *et al.*, 1994). Em vários autores esses dois tipos de fronteiras são chamados respectivamente de terminais e não terminais.

Mas alguns dos autores que consideram a distinção entre fronteiras terminais e não terminais defendem que existe também uma diferença interna. Para esses autores, não haveria um único tipo de fronteira terminal e um único tipo de fronteira não terminal. Segundo essa proposta, podemos observar terminais “mais terminais” que as outras. Por exemplo, as fronteiras de enunciados seriam menos terminais se comparados à fronteira entre blocos discursivos maiores, chamados de parágrafos por alguns (van DONZEL, 1999). Analogamente, existiriam diversos tipos de fronteiras não terminais, algumas mais salientes que outras, ou perceptualmente mais próximas das terminais, ou que anunciam o fato que a conclusão está próxima; essas propostas não devem ser consideradas excludentes mas podem capturar diferentes aspectos

da complexidade do fenômeno (SWERTS *et al.*, 1994; TEIXEIRA FALCÃO, 2017).

De fato, se examinarmos os parâmetros fonético-acústicos correlacionados à percepção de fronteira, em particular de fronteira não terminal, observamos combinações muito variadas dentro da mesma língua e do mesmo texto (cf. TEIXEIRA FALCÃO, 2017). Temos, por exemplo, fronteiras marcadas claramente por um movimento de f_0 ascendente, uma pista acústica de continuidade, que, frequentemente junto com outras marcas prosódicas como a duração, veicula a clara percepção de que o discurso vai continuar. Por outro lado, esse movimento ascendente de f_0 ou o alongamento final podem faltar em outras fronteiras que também são percebidas como não terminais (cf. WAGNER, 2010).

Quanto às fronteiras conclusivas, frequentemente se observa que elas são caracterizadas por um movimento descendente da curva de f_0 até o nível mais baixo, e seguidas por um reset da f_0 no começo da unidade seguinte, que iniciaria com uma f_0 em uma altura claramente distinta. Contudo é comumente reconhecido que nem todos os enunciados se concluem com uma f_0 baixa. Embora o caso mais evidente e estudado seja aquele das interrogativas polares em línguas como inglês e espanhol peninsular, há outras ilocuções, segundo a terminologia e categorização que adotamos, que são marcadas, entre outros parâmetros por uma f_0 final mais alta (CRESTI, 2000 e no prelo; MORAES; RILLIARD, 2014, *inter alia*).

A variabilidade na manifestação física das fronteiras pode estar correlacionada com diferentes valores funcionais no plano linguístico. Teríamos então não somente uma correlação entre tipos de fronteiras que veiculam conclusão e tipos de fronteiras que veiculam continuação, mas também entre tipos conclusivos diferentes, por exemplo de ilocuções diferentes, e entre tipos não conclusivos diferentes, que, por hipótese, marcariam tipos de constituintes diferentes (sintáticos ou de outra natureza). Nesse caso a manifestação específica de uma fronteira prosódica não teria apenas um valor demarcativo, mas dependeria fortemente da função linguística da unidade da qual marcam a fronteira e, portanto, pistas que assinalam também essas mesmas funções linguísticas.

Olhando por esse lado, estudar a forma como fisicamente as fronteiras são realizadas e estudar a natureza das unidades demarcadas por duas fronteiras (uma à esquerda e outra à direita) não seriam mais de âmbitos distintos, o primeiro de interesse prioritário da fonética e o segundo de níveis linguísticos superiores ou dos estudiosos dos mecanismos

cognitivos, mas se tornariam bem mais integrados. Essa perspectiva que une as funções das unidades à manifestação concreta das fronteiras que as delimitam é ainda incipiente e poderá nos dar respostas interessantes sobre a natureza das unidades que são delimitadas por fronteiras.

Antes de passar às diferentes abordagens teóricas sobre as unidades, vale a pena fazer uma observação sobre alguns tipos de fronteiras (e de unidades) que estão bem menos frequentes na fala de laboratório, ao menos quando essa se restringe à fala lida, mas que são extremamente comuns na fala espontânea: os diferentes tipos de disfluências. Na fala espontânea são muito frequentes os fenômenos de interrupção, de retratação e hesitação. Muitas unidades terminam não porque o falante planejou a sua conclusão, mas porque alguma causa imprevista de natureza interna (não recuperação de palavra apropriada, mudança de idéia, ou qualquer problema na articulação ou na elaboração do conteúdo) ou externa (interrupção por outro falante ou qualquer evento ambiental) leva à interrupção momentânea do enunciado antes que ele seja completado semântica e prosodicamente. Quanto à retratação, o enunciado não é interrompido, mas é fragmentado por causa de repetições de palavras ou partes de palavras, que depois o falante idealmente cancela e corrige, prosseguindo no enunciado como se elas não tivessem sido pronunciadas. Trata-se nesse caso do resultado de dificuldades na realização do enunciado que não levam à interrupção do mesmo e que são mais ou menos presentes em todos os falantes, mas principalmente naqueles que tem menor domínio da fala, seja porque são muito jovens, seja porque são de diastratia baixa, seja por outras razões. No caso da hesitação, as dificuldades na fala se manifestam de diferentes formas como alongamentos vocálicos ou tomadas de tempo, também chamadas de pausas preenchidas. Sempre ou quase sempre que se dão um desses três fenômenos se geram também uma ou duas fronteiras (geralmente uma no caso da interrupção e duas nos outros dois casos). Contudo, essas fronteiras em princípio não são planejadas pelo falante e não marcam unidades com função linguística. Na análise das pistas de fronteira prosódica elas constituem um elemento de ruído, não podendo ser comparadas às fronteiras que o falante faz para construir o significado do enunciado.

Um último tipo de fronteira que deve ser considerado é o que delimita aquelas que, no modelo da *Language into Act Theory* (L-Act; CRESTI, 2000; MONEGLIA; RASO, 2014; MONEGLIA; CRESTI,

1997), são chamadas de *Scanning Units*. Uma *Scanning Unit*, segundo a visão da L-AcT é uma unidade informacionalmente não autônoma e que constitui uma parte de uma unidade informacional maior (por ex. um Tópico dividido em duas ou mais unidades entonacionais. Nesse caso, as unidades antes da última seriam *Scanning Units*, enquanto o perfil prosódico que marca a função da unidade informacional é sempre colocado na última unidade entonacional). Segundo a L-AcT, as fronteiras que delimitam essas unidades são devidas a diferentes possíveis causas: ênfase (para realçar as partes que compõem o texto de uma unidade informacional se segmenta seu conteúdo em mais unidades entonacionais), imperícia do falante (como se fossem pequenas hesitações ou retratações sem acréscimo de material segmental), necessidade articulatória (quando uma unidade informacional possui uma quantidade de sílabas superior àquela que cabe confortavelmente dentro de uma unidade entonacional). Essas fronteiras, que, como vimos, constituem um grupo não homogêneo, têm uma tipologização complexa em relação às outras fronteiras, pois só é possível individualizar uma *Scanning Unit* depois de uma etiquetagem informacional, o que segue a segmentação e não pode ser automatizado.

Além de todas essas questões abertas, seria interessante ainda considerar diversas outras questões de ordem não linguística: as vozes masculinas e as femininas usam os parâmetros acústicos para veicular a percepção de fronteira da mesma maneira? O que acontece nas várias patologias de fala, quando são comprometidas funções articulatórias ou cognitivas? E como evolui a capacidade de gerenciar os parâmetros para esse objetivo funcional ao longo da ontogênese?

A pesquisa das últimas décadas avançou muito na compreensão e na investigação das complexas combinações de fatores que condicionam a manifestação de fronteiras, e mais recentemente os trabalhos estão começando a investigar o fenômeno na fala espontânea. Contudo, ainda resta um longo caminho a ser percorrido. Enfim, enfrentar a questão das combinações de parâmetros não é suficiente, é necessário também olhar para o valor de cada parâmetro nas diversas combinações e para o peso relativo (hierarquia) desses parâmetros dentro de cada combinação. É evidente que isso aumenta muito as variáveis responsáveis para a marcação de fronteira prosódica, o que de fato nos impõe o uso de instrumentos computacionais e estatísticos para apreendê-las em algum grau satisfatório.

Mais recentemente, as fronteiras prosódicas têm sido objeto de investigação da psicolinguística no que tange questões de processamento (DRURY *et al.*, 2016; GLUSHKO *et al.*, 2016; NICKELS *et al.*, 2013; HWANG; STEINHAUER, 2011; PAUKER *et al.* 2011; STEINHAUSER, 2003; STEINHAUER; FRIEDERICI, 2001), de modo particular através da técnica do *Event-Related Potential* (ERP). Foi Steinhauer *et al.* (1999) e colaboradores que inicialmente usaram a técnica de ERP para mostrar que fronteiras prosódicas ouvidas estão associadas a trechos de aumento da amplitude da atividade elétrica (potencial evocado), que foi nomeado de CPS (*Closure Positive Shift*). Esse pico ocorre de 400 a 800 ms. após um momento definido, que, nos testes mais bem sucedidos, foi considerado a partir da última tônica antes de fronteira. Os experimentos foram realizados com e sem a presença de pausa e de diversos outros parâmetros considerados responsáveis por veicular percepção de fronteira, mas o pico de atividade se manteve sempre. Parece que a presença do alongamento silábico e de um tom de fronteira são suficientes para que o encéfalo do ouvinte reaja. As pesquisas atuais procuram refinar cada vez mais a observação de como reagimos a parâmetros isolados ou às suas combinações quanto à percepção de fronteira.

É especialmente interessante o fato de que a segmentação (*phrasing*) seria sensível a pistas de modalidades diferentes: não somente pistas acústicas, mas também pistas gráficas, como a vírgula na leitura, causariam um aumento de atividade elétrica em correspondência de fronteira. Além disso, o fenômeno aparece também para a segmentação musical, mas com uma latência maior (talvez devida à ausência de informações de natureza linguística como a sintaxe e o léxico). Parece também que o CPS é encontrado somente a partir de uma certa idade (cerca de três anos de idade), o que faz com que ele seja considerado dependente de uma capacidade de estruturação mínima, seja de natureza sintática seja mesmo de natureza prosódica *stricto sensu*. Este resultado parece compatível com os dados de estudos aquisicionais (THORNTON, 2016; HYAMS; ORFITELLI, 2015, *inter alia*). Por fim, o CPS parece ser maior quanto menos esperada for a fronteira, ou seja, quanto menos ela seja previsível com base em informações de outra natureza; mas também parece bastante claro que a prosódia, como veículo de fronteira, seria capaz de prevalecer em caso de conflito com as expectativas de natureza sintática (BÖGELS, TORREIRA, 2015; BÖGELS *et al.*, 2013, 2010).

Como a fronteira é marcada pelo concurso de todos os parâmetros prosódicos, especialmente, duração silábica, f0 e intensidade, é importante ainda apontar que há uma predominância, em indivíduos destros, de processamento temporal no hemisfério esquerdo enquanto o processamento espectral ativa majoritariamente áreas do hemisfério direito (ROBIN *et al.*, 1990; ZATORRE, 1997), o que também é confirmado por estudos em indivíduos lesionados seja no hemisfério esquerdo, seja no hemisfério direito, com os primeiros perdendo capacidade de processamento temporal (SHAH *et al.*, 2006). Quanto às áreas neuronais envolvidas na percepção da fala, tanto as áreas corticais temporais quanto parietais são ativadas bilateralmente (HICKOK; POEPEL, 2000).

3. Segmentação e significado linguístico

A segmentação da fala é de fundamental importância para a construção do significado linguístico (cf. FERY, 2017, para uma revisão). A prosódia é utilizada para orientar o ouvinte na reconstrução de unidades funcionais distintas e de sua hierarquia e função na decodificação da mensagem. Essa é a razão principal que motiva os pesquisadores a estudarem a natureza física das fronteiras e a sua relação com os diferentes níveis linguísticos. Olhemos apenas alguns exemplos em línguas diferentes. Em inglês, uma sequência como *People give John the book I promised him* pode ser segmentada pelo menos das quatro maneiras seguintes, gerando significados muito diferentes entre eles, tanto do ponto de vista ilocucionário quanto sintático:

- (a) *People* (Calling)! *Give John the book I promised him* (Order)!
- (b) *People give John the book I promised him* (Assertion).
- (c) *People give John the book* (Question)? *I promised him* (Assertion).
- (d) *People* (Calling)! *Give John the book* (Order)! *I promised him* (Assertion).

Em (a), (c) e (d) teríamos duas fronteiras terminais, enquanto em (b) teríamos apenas uma fronteira, essa também de natureza terminal. Quanto aos parâmetros acústicos, no entanto, as fronteiras terminais das várias segmentações são diferentes pelo menos quanto ao movimento de

f0. Se a segunda fronteira de (a), (c) e (d) é precedida de um movimento descendente, sua primeira fronteira apresenta um movimento ascendente. Esses movimentos ascendentes não são iguais, assim como não são iguais os movimentos descendentes dos outros casos. Uma diferenciação análoga poderia ser feita para os valores de intensidade e duração.

Em português uma sequência como *João vai pro Rio até amanhã* (*João will go (or go) to Rio until tomorrow (or see you tomorrow)*) pode ser segmentada pelo menos das três maneiras seguintes:

- (a) *João* (calling)! *Vai pro Rio até amanhã* (order)! (*João! Go to Rio until tomorrow*)
- (b) *João vai pro Rio até amanhã* (assertion). (*João will go to Rio until tomorrow*)
- (c) *João* (calling)! *Vai pro Rio* (order)! *Até amanhã* (greeting)! (*João! Go to Rio! See you tomorrow*)

Nessas três organizações de sentenças, é evidente que a segmentação afeta a interpretação sintática e semântico-pragmática da sequência.

Por fim, o exemplo seguinte mostra como a segmentação pode ser decisiva para a interpretação sintática e semântica em outra língua ainda, nesse caso o italiano:

- (a) *Claudia* (calling)! *Guarda* (deixis)! *Quanto è bello* (expressive)! (*Claudia! Look! How beautiful it is!*)
- (b) *Claudia* (calling)! *Guarda quanto è bello* (deixis)! (*Claudia! Look how beautiful it is!*)
- (c) *Claudia guarda quanto è bello* (assertion). (*Claudia looks how beautiful it is.*)

A exemplificação poderia ser mais complexa, levando em conta outras interpretações e diferentes tipos de unidades) e estendida a outras línguas, mas o que importa aqui evidenciar é a importância do papel da segmentação na construção do significado linguístico tanto no nível sintático quanto naquele semântico. A presença de fronteira afeta certamente também o nível morfo-fonológico, por exemplo, inibindo fenômenos de sândi.

Nos exemplos anteriores vimos somente casos de fronteiras terminais, que isolam sequências autônomas pragmática e prosodicamente e que podem ser enunciadas em isolamento. Mas o significado é afetado também em caso de fronteiras não terminais, ou seja, quando a relação sintática ou informacional entre as unidades separadas pela fronteira deve ser mantida. Por exemplo a sequência *the film I like it* é analisável como um sintagma nominal modificado por uma relativa. Se, por outro lado, inserirmos uma fronteira a análise pode mudar: *in the film, I like it* a análise pode ser aquela de Tópico e Comentário e interpretável como: *as for the film, I like it*.

Voltemos agora à noção de unidade de referência da fala, entendida como unidade com sentido comunicativo autônomo que compõe o texto. Se consideramos a dimensão prosódica, é difícil definir essa unidade com base exclusivamente em critérios sintáticos que definem as tradicionais categorias de oração e sentença. A prosódia tem uma dimensão comunicativa que conduz preferencialmente os pesquisadores a prestar atenção à produção e à percepção da fala, mesmo não faltando perspectivas mais abstratas (mas fora de um contexto comunicativo). Muitos linguistas que incorporam a prosódia como elemento primário de seus modelos consideram a percepção prosódica de conclusão de uma sequência comunicativa como a marca principal da unidade de referência (CRESTI, 2000; MONEGLIA; RASO, 2014; IZRE'EL, 2002). Outros preferem considerar como unidade de referência a unidade entonacional, independentemente de ela apresentar um contorno prosódico percebido como conclusivo ou continuativo (METTOUCHI *et al.*, 2010). Em ambas as perspectivas a marca principal da unidade de referência está na fronteira entonacional. A diferença reside na questão de se qualquer fronteira pode marcar o fim de uma unidade de referência ou se somente fronteiras com uma qualidade específica fazem isso. Essa discussão se acompanha também daquela relativa às relações linguísticas que se manifestam dentro da unidade entonacional, dentro de um conjunto de unidades entonacionais marcado por uma fronteira conclusiva, e também nas relações que atravessam a fronteira conclusiva e precisam de unidades ainda maiores (para alguns aspectos dessa discussão dentro de arcabouços teóricos diferentes embora próximos, veja-se Izre'el neste número; CRESTI, 2014; PIETRANDREA *et al.*, 2014).

4. As contribuições presentes no volume e as contribuições para o debate

Os nove trabalhos apresentados neste número temático tocam em aspectos diferentes da segmentação prosódica da fala espontânea. Um primeiro grupo de contribuições se concentra na elaboração de softwares que permitam a extração de dados e informações capazes de esclarecer algumas das tantas questões ligadas à segmentação. Naturalmente, também por trás desses trabalhos há sempre uma hipótese teórica, seja na função, seja na quantidade de fronteiras a serem identificadas.

O trabalho de Xu e Gao apresenta a ferramenta computacional FormantPro que usa o *software* Praat como plataforma para a extração automática de trajetórias de formantes. Embora o tema não enfoque diretamente a questão da segmentação prosódica, tanto a ferramenta quando os exemplos levantados pelos autores abrem uma discussão sobre isomorfismo entre eventos acústicos e articulatórios que marcam fronteiras de consoantes e vogais. Essas fronteiras são discutidas a partir de um alinhamento com trajetórias de f_0 que podem vir a ter implicações para a delimitação de fronteiras prosódicas. O programa gera subsidiariamente valores de duração e de intensidade e permite a apresentação das trajetórias médias aliadas a uma normalização temporal que auxilia a observar as equivalências entre instâncias de diferentes enunciados contendo palavras em contraste. Os valores de duração podem ser usados para investigar pistas de fronteiras prosódicas no caso de mudanças importantes em relação ao seu entorno.

O trabalho de Teixeira Falcão e Mittmann apresenta um interessante procedimento para extrair modelos de parâmetros acústicos para diferentes tipos de fronteira em trechos de *corpora* de fala espontânea previamente segmentados por 14 segmentadores. Dos dados de *corpora*, após deles serem tratados para que o *script* pudesse lê-los em Praat, é extraído um número muito grande de medidas em uma janela de 10 unidades V-V à esquerda e à direita de cada posição candidata a fronteira, ou seja, de cada fronteira de palavra fonológica. A segmentação em unidade V-V (BARBOSA, 2006) mostra como outros níveis de segmentação da fala interagem necessariamente com o nível da unidade entonacional. Um procedimento estatístico e o procedimento humano de refinamento revelam as combinações de parâmetros que melhor explicariam as fronteiras, e seus pesos. Todo o trabalho foi

planejado considerando que as fronteiras prosódicas podem ser divididas em dois grandes grupos: terminais e não terminais. O trabalho relativo às quebras não terminais aponta para a necessidade de considerar essas fronteiras em pelo menos três sub-grupos distintos, cada um explicado por um modelo distinto. Esse resultado alimenta as reflexões tanto sobre a existência de uma distinção entre fronteiras terminais e não terminais e também de distinções mais sutis. Seria importante investigar ao que seriam devidas essas últimas.

O trabalho de Bigi e Meunier avalia a ferramenta de *software* SPPAS que permite a segmentação automática da fala lida e da espontânea focando especialmente, no último caso, nas questões relativas às disfluências. A ferramenta apresentada pressupõe a existência de uma transcrição ortográfica e de um dicionário de pronúncia de palavras de um léxico. Além disso, contém um modelo acústico dos sons da fala do francês que permite o alinhamento de símbolos fonéticos com o sinal de fala. Os erros de alinhamento são cerca de 11% na fala lida e 15 % na fala espontânea, mas podem ser reduzidos a partir de uma transcrição ortográfica enriquecida que identifique os tipos de disfluência. A ferramenta é testada com nove *corpora* que incluem fala lida, conversa espontânea e debate político para os casos de trechos disfluents contendo risos, pausas preenchidas e ruídos. Os autores demonstram que, sendo precedido de um pré-processamento que separa a cadeia de fala em unidades entre pausas, pode-se atingir um nível de precisão na delimitação dessas unidades de apenas 20 ms.

O artigo de G. Christodoulides usa dois *corpora* de fala do francês com anotação de fronteiras de força diferente para verificar: (a) o grau de acordo entre anotações prosódicas a partir de duas abordagens teóricas diferentes, a teoria métrica autosegmental (PIERREHMBERT, 1980) e a distinção entre micro e macro-sintaxe (BLANCHE-BENVENISTE, 2002, 2003) quanto a dois níveis de anotação comparáveis; (b) quais parâmetros acústicos são mais importantes para veicular os dois tipos de fronteira e qual a sua hierarquia. O uso de *corpora* dependentes de abordagens teóricas tão diferentes é um teste importante para o tema das fronteiras, ainda mais considerando que um *corpus* é segmentado com base em critérios teóricos e o outro com base em critérios perceptuais. Os parâmetros investigados são a presença e a duração de pausa, o alongamento pré-fronteiriço e duas medidas de f_0 associadas com a fronteira. A análise mostra um acordo muito alto entre os dois *corpora*

quanto aos parâmetros prosódicos envolvidos nas posições onde foi marcada a fronteira e na distinção dos dois tipos de fronteiras comparados. A conclusão é que o parâmetro mais importante associado com a presença de fronteira e com a sua maior força seria a pausa, seguida pelo alongamento silábico. A f0 seria importante apenas para marcar a diferença entre presença ou ausência de fronteira, mas não para distinguir as forças dos dois tipos de fronteira.

O trabalho de Ph. Martin se distingue dos outros por analisar uma unidade diferente: o grupo acentual. O objeto de estudo é, portanto, uma unidade menor que a unidade entonacional, mesmo se às vezes pode coincidir com ela. A observação dessa unidade é importante para os nossos objetivos, já que Martin individualiza um número restrito de possíveis movimentos de f0 dentro de um grupo acentual e um número restrito de sequências de movimentos dentro da unidade entonacional, com um critério de dependência entre eles. Isso nos permite investigar a estrutura interna de uma unidade entonacional com base em unidades marcadas pelo acento. Entre outras consequências, os resultados dessa análise poderão trazer mais luz sobre as características das diversas estruturas internas das unidades entonacionais e sobre quanto e como essas estruturas se correlacionam com a função linguística veiculada pela unidade. Diferentes aspectos da unidade, com a presença de certas proeminências em certas posições, já são discutidos na literatura, mesmo se ainda não suficientemente na perspectiva adotada aqui. Propostas como a de Martin nos levam a considerar o papel de um outro nível prosódico e a sua específica função linguística, além de outras características (proeminências, tipo de fronteira) que possam nos fazer entender melhor como construímos uma sequência que possui uma função linguística gerenciando diferentes níveis da estrutura prosódica.

Um terceiro grupo de artigos investiga as fronteiras com objetivos diretamente ligados a algum nível linguístico, seja sintático, seja informacional.

O trabalho de A. Mettouchi, realizado sobre o Kabyle, língua afro-asiática da Argélia, mostra como a presença/ausência de fronteira pode constituir uma marca de uma função de natureza sintática, nesse caso o objeto direto. A fronteira se revela como o traço formal decisivo para distinguir essa estrutura de outras estruturas que possuem funções diferentes, talvez de natureza especificamente informacional, mas que aparecem no enunciado com os mesmos traços formais do objeto, a não

ser pela presença da fronteira que é ao contrário ausente na realização com função de objeto direto. Esse trabalho levanta uma questão importante: a relação entre presença de fronteira e ruptura da composicionalidade sintática. Outros estudos (CRESTI, 2014; RASO; VIEIRA, 2016) tratam dessa importante questão, que ainda é controversa. Se por um lado é fácil encontrar casos em que parece que a composicionalidade sintática se interrompe em coincidência da fronteira (e seria, portanto, possível hipotetizar que a fronteira ou algum tipo de fronteira tenha um papel em marcar essa interrupção), por outro lado outros casos são interpretáveis também salvando a composicionalidade em presença de fronteira.

O trabalho de da Silva e Fonseca apresenta também diversos motivos de interesse. Um primeiro motivo, como no caso do artigo anterior e do seguinte, é a importância que uma marca prosódica adquire para a identificação de uma unidade linguística, no caso em questão o Tópico. Um segundo motivo é a natureza experimental do trabalho, sobre a qual diremos algo mais para frente. Um terceiro motivo é que mostra como os resultados apresentados dentro de uma teoria formalista podem ser úteis também para diferentes visões da categoria de Tópico, evidenciando como o estudo empírico dos dados beneficia o debate. Os experimentos idealizados e realizados por da Silva e Fonseca podem ser muito interessantes para o debate entre os estudiosos da estruturação informacional da fala. Seus resultados podem ser utilizados para comparar a definição sintática de Tópico com as definições mais pragmáticas, em particular com a definição proposta por L-AcT, que dá um peso muito grande aos aspectos prosódicos, e que apresenta vários resultados de pesquisas em diversas línguas, entre as quais o PB (cf. CRESTI, 2000; SIGNORINI, 2004; FIRENZUOLI; SIGNORINI, 2003; MONEGLIA; RASO 2014; ROCHA; RASO, 2013; CAVALCANTE, 2016; MITTMANN, 2012; RASO; CAVALCANTE; MITTMANN, no prelo). De fato, os resultados não esperados do terceiro experimento poderiam ser explicados assumindo que o Tópico seja uma categoria pragmática que não depende da estrutura argumental e, portanto, pode estar em posição de sujeito, mas ser marcada por fronteira e por um foco prosódico funcional que o distingue claramente do sujeito, que por sua vez não apresenta fronteira prosódica com o resto do enunciado e não possui foco prosódico funcional. Nesse caso, a diferença entre sujeito e Tópico não seria de natureza sintática, mas seria devida à diferença entre um item sintático interno ao comentário (o sujeito) e um item pragmático

externo ao comentário (o Tópico). Um diálogo mais aprofundado entre essas diferentes abordagens teóricas poderia levar a uma maior clareza sobre o conceito de Tópico e estimular ambas as teorias a refinar a própria análise e as próprias argumentações, utilizando tanto abordagens experimentais como aquelas propostas por Da Silva e Fonseca quanto os dados de *corpora* de fala espontânea elaborados dentro de L-AcT.

Também o artigo de Panunzi e Saccone é fortemente orientado teoricamente. De fato, seu objetivo é observar se, quanto e como as fronteiras entre pares de unidades informacionais se realizam com características diferentes entre elas. Os dois pares (ou raramente sequências de mais de duas unidades) que são explorados no artigo são diferentes combinações de unidades ilocucionárias. Um tipo de par é caracterizado por compor uma interpretação ilocucionária complexa mas única, prosódica e pragmaticamente padronizada. O outro tipo de par, ao contrário, é constituído por duas ilocuções independentes, apesar de serem separadas por fronteira não terminal. Portanto, para analisar as fronteiras, é necessário que o texto já esteja anotado informacionalmente segundo um determinado arcabouço teórico, neste caso o da L-AcT (CRESTI, 2000; MONEGLIA; RASO, 2014). Os primeiros resultados parecem apontar para diferenças fortes na forma das fronteiras entre os dois pares de unidades. Este é um exemplo instigante de como as características das fronteiras podem correlacionar com a função das unidades que elas separam. Esse tipo de trabalho, que tenta correlacionar função da unidade com as características das fronteiras pode ser aplicado a diferentes tipos de unidades e com base em diferentes quadros teóricos.

O trabalho de Izre'el foi colocado no final deste número porque, a partir da consideração dos aspectos prosódicos e em particular das fronteiras, propõe uma revisão geral das categorias tradicionais de frase, oração, sentença e predicação, mostrando como a incorporação dos traços prosódicos pode levar a uma reformulação geral das categorias canônicas no estudo da fala espontânea. Izre'el retoma a discussão linguística sobre essas categorias desde a linguística clássica até Chomsky, para mostrar como certas categorias, assim como definidas na tradição de imposição sintaticista, não funcionam na análise da fala e principalmente da fala espontânea que, em princípio, deve ser o domínio natural para análise da linguagem. Levando em conta a prosódia e dados de *corpora* de fala espontânea, emerge claramente a importância da categoria de ilocução (que Izre'el chama de *modalidade*) enquanto categoria decisiva para a

individualização de uma unidade comunicativa e enquanto categoria marcada diretamente pela prosódia. E ainda emerge claramente a importância das fronteiras prosódicas para definir o domínio em que se dão as relações linguísticas em sua realização comunicativa. Como outros trabalhos deste número, mas com um escopo maior, a contribuição de Izre'el traz mais argumentos (cf. também BIBER *et al.*, 1999; as contribuições em RASO; MELLO, 2014; CRESTI, 2005; RASO; MITTMANN, 2012, *inter alia*) que apontam para a necessidade de definir a unidade comunicativa da fala revendo a noção de predicação (e de proposição), ou as noções de oração e de sentença, e sustentando a necessidade de incorporar a prosódia como elemento central na marcação dessa unidade de referência comunicativa. Como diversas contribuições deste número, o artigo de Izre'el não deixa dúvida sobre a necessidade de incorporar a prosódia entre os níveis de análise da linguística, e, mais do que isso, sobre o peso hierarquicamente decisivo da prosódia na individualização dos constituintes linguísticos da fala.

Referências

- AMIR, N.; SILBER-VAROD, V.; IZRE'EL, S. Characteristics of Intonation Unit Boundaries in Spontaneous Spoken Hebrew: Perception and Acoustic Correlates. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 2004. Nara. *Proceedings...* Nara: ISCA, 2004. p. 677-680.
- AURAN, C.; BOUZON, C.; HIRST, D. The Aix-MARSEC Project: an Evolutive Database of Spoken British English. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 2004. Nara, Japan. *Proceedings...* Nara: ISCA, 2004.
- AVANZI, M.; LACHERET-DUJOUR, A.; VICTORRI, B. ANALOR. Tool for Semi-Automatic Annotation of French Prosodic Structure. In: ANALOR. A Tool for Semi-Automatic Annotation of French Prosodic Structure. Campinas, Brazil, May 2008. p. 119-122.
- BARBOSA, P. A. *Caractérisation et génération automatique de la structuration rythmique du français*. 1994. Tese (Doutorado) – Institut National Polytechnique de Grenoble, França, 1994.
- BARBOSA, P. A. *Incurções em torno do ritmo da fala*. Campinas: Pontes, 2006.

BARBOSA, P. A. Prominence-and Boundary-Related Acoustic Correlations in Brazilian Portuguese Read and Spontaneous Speech. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 4., 2008, Campinas. *Proceedings...* Campinas: ISCA, 2008. p. 257-260.

BARBOSA, P. A. Automatic Duration-Related Saliency Detection in Brazilian Portuguese Read and Spontaneous Speech. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 5., 2010, Chicago. *Proceedings...* Chicago: ISCA, 2010.

BARBOSA, P. A. Panorama of Experimental Prosody Research. In: GSCP INTERNATIONAL CONFERENCE – SPEECH AND CORPORA, VIIth., Belo Horizonte. *Proceedings...* Florence: Firenze University Press, 2012. p. 33-42.

BARTH-WEINGARTEN, D. *Intonation Units Revisited*. Cesura in Talk-In-Interaction. Amsterdam: John Benjamins, 2016.

BIBER, D.; JOHANSSON, S.; LEECH, G.; CONRAD, S.; FINEGAN, E. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited, 1999.

BLANCHE-BENVENISTE, C. Macro-syntaxe et micro-syntaxe: les dispositifs de la rection verbale. In: ANDERSEN, H. L.; NØLKE, H. (Éd.). *Macro-Syntaxe e Macro-Sémantique*. Bern: Peter Lang, 2002. p. 95-115.

BLANCHE-BENVENISTE, C. Le recouvrement de la syntaxe et de la macro-syntaxe. In: SCARANO, A. (Ed.). *Macro-syntaxe et pragmatique*. L'analyse linguistique de l'oral. Roma: Bulzoni, 2003. p. 53-75.

BLANCHE-BENVENISTE, C.; JEANJEAN, C. *Le français parlé*. Transcription et édition. Paris: Didier Érudition; Institut National de la Langue Française, 1987.

BLANCHE-BENVENISTE C. Macro-syntaxe et micro-syntaxe: les dispositifs de la rection verbale. In: ANDERSEN, H. L.; NØLKE, H. (Éd.). *Macro-syntaxe et macro-sémantique*. Actes du Colloque International d'Århus [17-19 mai 2001]. Bern: Peter Lang, 2002. p. 95-118.

BÖGELS, S.; SCHRIEFERS, H.; VONK, W.; CHWILLA, D. J.; KERKHOFS, R. The Interplay Between Prosody and Syntax in Sentence Processing: The Case of Subject- and Object-Control Verbs. *Journal of Cognitive Neuroscience*, Cambridge, v. 22, n. 5, p. 1036-1053, 2010.

- BÖGELS, S.; SCHRIEFERS, H.; VONK, W.; CHWILLA, D.; KERKHOF, R. Processing Consequences of Superfluous and Missing Prosodic Breaks in Auditory Sentence Comprehension. *Neuropsychologia*, Oxford, v. 51, p. 2715-2728, 2013.
- BÖGELS, S.; TORREIRA, F. Listeners Use Intonational Phrase Boundaries to Project Turn Ends in Spoken Interaction. *Journal of Phonetics*, [s.l.], v. 52, p. 46-57, 2015.
- BOLINGER, D. Pitch Accent and Sentence Rhythm. In: ABE, I.; KANEKIYO, T. (Ed.). *Forms of English: Accent, Morpheme, Order*. Cambridge, Mass: Harvard University Press, 1965. p. 139-180.
- BOSSAGLIA, G.; MELLO, H.; RASO, T. Insubordination and the Syntax/Prosody Interface in Spoken Brazilian Portuguese: Data on Adverbial Clauses. In: IZRE'EL, S.; MELLO, H.; PANUNZI, A.; RASO, T. *In Search for a Reference Unit of Spoken Language: A Corpus Driven Approach*. Amsterdam: John Benjamins. (Forthcoming).
- BUHMANN, J.; CASPERS, J.; HEUVEN, V. J. van; HOEKSTRA, H.; MARTENS, J-P.; SWERTS, M. Annotation of Prominent Words, Prosodic Boundaries and Segmental Lengthening by Non-Expert Transcribers in the Spoken Dutch Corpus. In: LREC, 3rd, 2002, Las Palmas. *Proceedings...* Las Palmas: ELRA, 2002. p. 779-785.
- BYBEE, J. *Language, Usage and Cognition*. Cambridge: CUP, 2010.
- CARLSON, R.; HIRSCHBERG, J.; SWERTS, M. Cues to Upcoming Swedish Prosodic Boundaries: Subjective Judgment Studies and Acoustic Correlates. *Speech Communication*, [s.l.], v. 46, p. 326-333, 2005. Doi: <https://doi.org/10.1016/j.specom.2005.02.013>
- CAVALCANTE, F. *The Topic Unit in Spontaneous American English: a Corpus-Based Study*. 2016. Master (Thesis) – Universidade Federal de Minas Gerais, Belo Horizonte, 2016.
- CHAFE, W. *Discourse, Consciousness and Time. The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: Chicago University Press, 1994.
- COUPER-KUHLEN, E. Prosody and sequence organization in English conversation. In: COUPER-KUHLEN, E.; FORD, C. E. (Ed.). *Sound Patterns in Interaction: Cross-Linguistic Studies from Conversation*. Amsterdam: John Benjamins, 2004. p. 335-376.

COWAN, N. *Attention and Memory: An Integrated Framework*. Oxford: Oxford University Press, 1998. Doi: <https://doi.org/10.1093/acprof:oso/9780195119107.001.0001>

CRESTI, E. *Corpus di italiano parlato*. Firenze: Accademia della Crusca, 2000. 2 v.

CRESTI, E. Notes on Lexical Strategy, Structural Strategies and Surface Clause Indexes in the C-ORAL-ROM Spoken Corpora. In: CRESTI, E.; MONEGLIA, M. (Ed.). *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam; Philadelphia: John Benjamins, 2005. p. 209-256.

CRESTI, E. Syntactic Properties of Spontaneous Speech in the Language into Act Theory: Data on Italian Complements and Relative Clauses. In: RASO, T.; MELLO, H. (Ed.). *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins, 2014.

CRESTI, E. The Pragmatic Analysis of Speech and Its Illocutionary Classification According to Language into Act Theory. In: IZRE'EL, S.; MELLO, H.; PANUNZI, A.; RASO, T. *In Search for a Reference Unit of Spoken Language: A Corpus Driven Approach*. Amsterdam: John Benjamins. (Forthcoming).

CRESTI, E.; MONEGLIA, M. (Ed.). *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: John Benjamins, 2005.

CROFT, W. Intonational Units and Grammatical Structure. *Linguistics*, [s.l.], v. 33, n. 5, p. 839-882, 1995.

CRUTTENDEN, Alan. *Intonation*. 2nd edition. New York: Cambridge University Press, 1997.

CRYSTAL, D. *Prosodic Systems and Intonation in English*. Cambridge: CUP, 1969.

DILLEY, L.; SHATTUCK-HUFNAGEL, S.; OSTENDORF, M. Glottalization of Word-Initial Vowel as a Function of Prosodic Structure. *Journal of Phonetics*, [s.l.], v. 24, p. 423- 444, 1996.

- DRURY, J. E.; BAUM, Sh. R.; VALERIOTE, H.; STEINHAUSER, K. Punctuation and Implicit Prosody in Silent Reading: An ERP Study Investigating English Garden-Path Sentences. *Frontiers in Psychology*, [s.l.], Sept. 2016. Doi: <https://doi.org/10.3389/fpsyg.2016.01375>
- DU BOIS, J. W.; CHAFE, W. L.; MEYER, Ch.; THOMPSON, S. *Santa Barbara Corpus of Spoken American English*. Washington DC: Linguistic Data Consortium, 2000-2005.
- DU BOIS, J. W.; CUMMING, S.; SCHUETZE-COBURN, S.; PAOLINO, D. Discourse transcription. *Santa Barbara Papers in Linguistics*, Santa Barbara, v. 4, n. 1, p. 225, 1992.
- EVANS, N.; WATANABE, H. The dynamics of insubordination: An overview. In: _____. (Ed.). *Insubordination*. Amsterdam: John Benjamins, 2016.
- FRY, C. *Intonation and Prosodic Structure*. Cambridge: CUP, 2017.
- 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.
- FON, J.; JOHNSON, K.; CHEN, S. Durational Patterning at Syntactic and Discourse Boundaries in Mandarin Spontaneous Speech. *Language and Speech*, [s.l.], v. 54, n. 1, p. 5-32, 2011.
- FUCHS, S.; KRIVOKAPIĆ, J.; JANNEDY, S. Prosodic Boundaries in German: Final Lengthening in Spontaneous Speech. *The Journal of the Acoustical Society of America*, [s.l.], v. 127, n. 3, p. 1851, 2010. Doi: 10.1121/1.3384378
- FURHT, B.; VILLANUSTRE, F. *Big Data Technologies and Applications*. [s.l.]: Springer, 2016.
- GAROFALO, J. S.; LAMEL, L. F.; FISHER, W. M.; FISCUS, J. G.; PALLETT, D. S.; DAHLGREN, N. L.; ZUE, V. TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1. Web Download. Philadelphia: Linguistic Data Consortium, 1993.
- GEE, J. P.; GROSJEAN, F. Performance Structures: A Psycholinguistic and Linguistic Appraisal. *Cognitive Psychology*, [s.l.], v. 15, n. 4, p. 411-458, 1983.

GLUSHKO, A.; STEINHAEUER, K.; De PRIEST, J.; KOELSCH, S. Neurophysiological Correlates of Musical and Prosodic Phrasing: Shared Processing Mechanisms and Effects of Musical Expertise. *PLoS ONE*, San Francisco, v. 11, n. 5, 2016.

GORDON, M.; LADEFOGED, P. Phonation Types: a Cross-Linguistic Overview. *Journal of Phonetics*, [s.l.], v. 29, p. 383-406, 2001.

GROSJEAN, F.; DOMMERGUES, J. Y. Les structures de performance en psycholinguistique. *L'Année Psychologique*, [s.l.], v. 83, n. 2, p. 513-536, 1983.

GROSJEAN, F.; GROSJEAN, L.; LANE, H. The Patterns of Silence: Performance Structures in Sentence Production. *Cognitive Psychology*, [s.l.], v. 11, n. 1, p. 58-81, 1979.

HANSON, H. M.; CHUANG, E. S. Glottal Characteristics of Male Speakers: Acoustic Correlates and Comparison with Female Data. *Journal of the Acoustical Society of America*, [s.l.], v. 106, n. 2, p. 1064-1077, 2001.

t'HART, J. Differential Sensitivity to Pitch Distance, Particularly in Speech. *The Journal of the Acoustical Society of America*, [s.l.], v. 69, n. 3, p. 811-821, 1981.

t'HART, J.; COLLIER, R.; COHEN, A. *A Perceptual Study on Intonation: An Experimental Approach to Speech Melody*. Cambridge: CUP, 1990. Doi: <https://doi.org/10.1017/CBO9780511627743>

HICKOK, G.; POEPPPEL, D. Towards a Functional Neuroanatomy of Speech Perception. *Trends in Cognitive Sciences*, [s.l.], v. 4, n. 4, p. 131-138, 2000.

HIRST, D.; Di CRISTO, A. A Survey of Intonation Systems. In: _____. (Ed.). *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press, 1998.

HUGGINS, A. W. F. Just Noticeable Differences for Segment Duration in Natural Speech. *The Journal of the Acoustical Society of America*, [s.l.], v. 51, n. 4B, p. 1270-1278, 1972.

HWANG, H.; STEINHAEUER, K. Phrase Length Matters: The Interplay Between Implicit Prosody and Syntax in Korean ‘Garden Path’ Sentences. *Journal of Cognitive Neuroscience*, Cambridge, v. 23, n. 11, p. 3555-3575, 2011. Doi: https://doi.org/10.1162/jocn_a_00001

HYAMS, N.; ORFITELLI, R. The Acquisition of Syntax. In: CAIRNS, H.; FERNANDEZ, E. (Ed.). *Handbook of Psycholinguistics*. [s.l.]: Wiley; Blackwell Publishers, 2015.

IZRE’EL, S. The Corpus of Spoken Israeli Hebrew: Textual Samples. *Lesson Enu*, v. 64, p. 289-314, 2002.

IZRE’EL, S.; MELLO, H.; PANUNZI, A.; RASO, T. (Ed.). *In Search for a Reference Unit of Spoken Language: A Corpus Driven Approach*. Amsterdam: John Benjamins. (Forthcoming)

KLATT, D. H.; COOPER, W. E. Perception of Segment Duration in Sentence Contexts. In: COHEN, A.; NOOTEBOOM, S. G. (Ed.). *Structure and Process in Speech Perception*. Berlin; Heidelberg: Springer, 1975. p. 69-89. Doi: https://doi.org/10.1007/978-3-642-81000-8_5

KOFFI, E. A Just Noticeable Difference (JND) Reanalysis of Fry’s Original Acoustic Correlates of Stress in American English. *Linguistic Portfolios*, St. Cloud, v. 7, 2018.

LADD, D. R. *Intonational phonology*. Cambridge: Cambridge University Press, 1996.

LIEBERMAN, P. Some Acoustic Correlates of Word Stress in American English. *Journal of the Acoustical Society of America*, [s.l.], v. 32, p. 451-454, 1960.

LINELL, P. *The Written Language Bias in Linguistics: Its Nature, Origins and Transformations*. London; New York: Routledge, 2005. Doi: <https://doi.org/10.4324/9780203342763>

MacNEILAGE, P. F. The Frame/Content Theory of Evolution of Speech Production. *Behavioral and Brain Sciences*, Cambridge, v. 21, n. 4, p. 499-511, 1998.

MARTIN, P. Les problèmes de l’intonation: recherches et applications. *Langue Française*, [s.l.], v. 19, p. 4-32, 1973.

MARTIN, P. Prosodic and Rhythmic Structures in French. *Linguistics*, [s.l.], v. 25, n. 5, p. 925-950, 1987.

MELLO, H.; RASO, T.; MITTMANN, M. VALE, H.; CÔRTEZ, P. Transcrição e segmentação prosódica do *corpus* c-oral-brasil: critérios de implementação e validação. In: RASO, T.; MELLO, H. (Ed.). C-ORAL-BRASIL I. *Corpus* de referência do português brasileiro falado informal. Belo Horizonte: UFMG, 2012. p. 125-174.

METTOUCHI, A.; CAUBET, D.; VANHOVE, M.; TOSCO, M.; COMRIE, B.; IZRE'EL, S. CORPAFROAS, A Corpus for Spoken Afroasiatic Languages: Morphosyntactic and Prosodic Analysis. In: ITALIAN MEETING OF AFRO-ASIATIC LINGUISTICS, 13th, 2010, Padova. *Proceedings...* Padova: Sargon, 2010. p.177-180.

MITTMANN, M. M. *O C-ORAL-BRASIL e o estudo da fala informal: um novo olhar sobre o Tópico no Português Brasileiro*. 2012. Dissertation (Ph.D.) – Universidade Federal de Minas Gerais, Belo Horizonte, 2012.

MITTMANN, M. M.; BARBOSA, P. A. An Automatic Speech Segmentation Tool Based on Multiple Acoustic Parameters. *CHIMERA. Romance Corpora and Linguistic Studies*, Madrid, v. 3, p.133-147, 2016.

MO, Y. Duration and Intensity as Perceptual Cues for Naïve Listeners' Prominence and Boundary Perception. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 4th., 2008, Campinas. *Proceedings...* Campinas: ISCA, 2008. p. 739-742.

MO, Y.; COLE, J.; LEE, E. K. Naïve Listeners' Prominence and Boundary Perception. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 4th., 2008, Campinas. *Proceedings...* Campinas: ISCA, 2008. p. 735-738.

MO, Y.; COLE, J. Perception of Prosodic Boundaries in Spontaneous Speech with and Without Silent Pauses. *The Journal of the Acoustical Society of America* [s.l.], v. 127, n. 3, p. 1956, 2010.

MONEGLIA, M.; CRESTI, E. L'intonazione e i criteri di trascrizione del parlato adulto e infantile. In: BORTOLINI, U.; PIZZUTO, E. (Ed.). *Il Progetto CHILDES Italia*. Pisa: Del Cerro, 1997. p. 57-90.

MONEGLIA, M.; FABBRI, M.; QUAZZA, S.; ANDREA; PANIZZA, A.; DANIELI, M.; GARRIDO, J. M.; SWERTS, M. Evaluation of Consensus on the Annotation of Terminal and Non-Terminal Prosodic Breaks in the C-ORAL-ROM *corpus*. In: CRESTI, E.; MONEGLIA, M. (Ed.). *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance*

Languages. Amsterdam: John Benjamins, 2005. p. 257-276. Doi: <https://doi.org/10.1075/scl.15.09mon>

MONEGLIA, M.; RASO, T. Notes Language into Act Theory (L-Act). In: RASO, T.; MELLO, H. (Ed.). *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins, 2014. p. 468-495. Doi: <https://doi.org/10.1075/scl.61.15mon>

MONNIN, P.; GROSJEAN, F. Les structures de performance en français: caractérisation et prédiction. *L'Année Psychologique*, [s.l.], v. 93, p. 9-30, 1993.

MORAES, J. A.; RILLIARD, A. Illocution, Attitude and Prosody: A Multimodal Analysis. In: RASO, T.; MELLO, H. (Ed.). *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins, 2014. p. 233-270. Doi: <https://doi.org/10.1075/scl.61.09mor>

NESPOR, Marina; VOGEL, Irene. *Prosodic Phonology*. Dordrecht: Foris Publications, 1986.

NI, C. J.; ZHANG, A. Y.; LIU, W. J.; XU, B. Automatic Prosodic Break Detection and Feature Analysis. *Journal of Computer Science and Technology*, [s.l.], v. 27, n. 6, p. 1184-1196, 2012.

NICKELS, S.; OPITZ, B.; STEINHAEUER, K. ERPs Show that Classroom-Instructed Late Second Language Learners Rely on the Same Prosodic Cues in Syntactic Parsing as Native Speakers. *Neuroscience Letters*, [s.l.], v. 557, p. 107-111, 2013.

OSTENDORF, Mari; PRICE, Patti; SHATTUCK-HUFNAGEL, Stefanie. Boston University Radio Speech Corpus LDC96S36. DVD. Philadelphia: Linguistic Data Consortium, 1996.

PAUKER, E.; ITZHAK, I.; BAUM, S. R.; STEINHAEUER, K. Effects of Cooperating and Conflicting Prosody in Spoken English Garden Path Sentences: ERP Evidence for the Boundary Deletion Hypothesis. *Journal of Cognitive Neuroscience*, Cambridge, v. 23, n. 10, p. 2731-2751, 2011. Doi: <https://doi.org/10.1162/jocn.2011.21610>

PIERREHUMBERT, J. B. *The Phonology and Phonetics of English Intonation*. 1980. Dissertation (Doctoral) – Massachusetts Institute of Technology, 1980.

PIETRANDREA, P.; KAHANE, S.; LACHERET, A.; SABIO, F. In: RASO, T.; MELLO, H. (Ed.). *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins, 2014.

PIKE, K. *The Intonation of American English*. Ann Arbor: University of Michigan Press, 1945.

PRICE, P. J.; OSTENDORF, M.; SHATTUCK-HUFNAGEL, S.; FONG, C. The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America*, [s.l.], v. 90, n. 6, p. 2956-2970, 1991.

RASO, T.; CAVALCANTE, F.; MITTMANN, M. M. Prosodic Forms of the Topic Information Unit in a Cross-Linguistic Perspective: a First Survey. In: GSCP INTERNATIONAL CONFERENCE, 2016, Napole. *Proceedings...* Napole. (Forthcoming).

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

RASO, T.; MELLO, H. (Ed.). *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins, 2014.

RASO, T.; MELLO, H. (Ed.). The C-ORAL-BRASIL II. *Corpus de referência do português falado (formal em contexto natural, mídia e telefone)*. (Em preparação).

RASO, T.; MITTMANN, M. As principais medidas da fala. In: RASO, T.; MELLO, H. (Ed.). C-ORAL-BRASIL I. *Corpus de referência do português brasileiro falado informal*. Belo Horizonte: UFMG, 2012. p. 177-220.

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*, Madrid, v. 3, p. 221-249, 2016.

REDI, L.; SHATTUCK-HUFNAGEL, S. Variation in the Realization of Glottalization in Normal Speakers. *Journal of Phonetics*, [s.l.], v. 29, p. 407-29, 2001.

RIETVELD, A.; GUSSENHOVEN, C. On the Relation Between Pitch Excursion Size and Prominence. *Journal of Phonetics*, [s.l.], v. 13, p. 299-308, 1985.

ROBIN, D. A. *et al.* Auditory Perception of Temporal and Spectral Events in Patients with Focal Left and Right Cerebral Lesions. *Brain and Language*, [s.l.], v. 39, p. 539-555, 1990.

ROCHA, B.; RASO, T. O pronome lembrete e a Teoria da Língua em Ato: uma análise baseada em corpora. *Veredas*, Juiz de Fora, v. 17, n. 2, p. 39-59, 2013.

SCHEGLOFF, E. A. Reflections on Studying Prosody in Talk-in-Interaction. *Language and Speech*, [s.l.], v. 41, n. 3-4, p. 235-263, 1998.

SCHUURMAN, I.; SCHOUPPE, M.; HOEKSTRA, H.; Van der WOUDE, T. CGN, an Annotated Corpus of Spoken Dutch. In: INTERNATIONAL WORKSHOP ON LINGUISTICALLY INTERPRETED CORPORA (LINC-03), 4TH, 2003, Budapest. *Proceedings...* Budapest: EACL, 2003.

SELKIRK, E. Sentence prosody: Intonation, Stress and Phrasing. In: GOLDSMITH, J. A. (Ed.). *The Handbook of Phonological Theory*. Oxford: Blackwell, 1995. p. 550-569.

SHAH, A, P.; BAUM, S. R.; DWIVEDI, V. D. Neural Substrates of Linguistic Prosody: Evidence from Syntactic Disambiguation in the Productions of Brain-Damaged Patients. *Brain and Language*, [s.l.], v. 96, p. 78-89, 2006.

SHRIBERG, E.; STOLCKE, A.; HAKKANI-TÜR, D.; TÜR, G. Prosody-Based Automatic Segmentation of Speech into Sentences and Topics. *Speech Communication*, [s.l.], v. 32, n. 1-2, p. 127-154, 2000.

SIGNORINI, S. *Topic e soggetto in corpora di italiano parlato spontaneo*. 2004. Dissertation (Ph.D.) – Università Firenze, Firenze, 2004.

STEINHAUER, K.; ALTER, K.; FRIEDERICI, A. D. Brain Potentials Indicate Immediate Use of Prosodic Cues in Natural Speech Processing. *Nature Neuroscience*, [s.l.], v. 2, n. 2, p. 191-196, 1999.

STEINHAUER, K. Electrophysiological Correlates of Prosody and Punctuation. *Brain and Language*, [s.l.], v. 86, n. 1, Special issue on the Neuronal Basis of Language, p. 142-164, 2003.

STEINHAUER, K.; FRIEDERICI, A. D. Prosodic Boundaries, Comma Rules, and Brain Responses: The Closure Positive Shift in ERPs as a Universal Marker for Prosodic Phrasing in Listeners and Readers. *Journal of Psycholinguistic Research*, [s.l.], v. 30, n. 3, p. 267-295, 2001.

SWERTS, M. Prosodic Features of Discourse Units. 1994. Thesis (PhD) – Technische Universiteit Eindhoven, 1994.

SWERTS, M. Prosodic Features at Discourse Boundaries of Different Strength. *The Journal of the Acoustical Society of America*, [s.l.], v. 101, n. 1, p. 514-521, 1997.

SWERTS, M.; COLLIER, R.; TERKEN, J. Prosodic Predictors of Discourse Finality in Spontaneous Monologues. *Speech Communication*, [s.l.], v. 15, n. 1-2, p. 79-90, 1994.

TEIXEIRA FALCÃO, B. H. *Correlatos fonético-acústicos de fronteiras prosódicas na fala espontânea*, 2017. Tese (Doutorado) – Universidade Federal de Minas Gerais, Belo Horizonte, 2017.

TESNIÈRE, L. *Éléments de Syntaxe Structurale*. Paris: C. Klincksieck, 1965.

THORNTON, R. Children's Acquisition of Syntactic Knowledge. In: ARONOFF, M. (Ed.). *Oxford Research Encyclopedia of Linguistics*, 2016. Doi: <https://doi.org/10.1093/acrefore/9780199384655.013.72>

TSENG, C. Y.; CHANG, C. H. Pause or No Pause? Prosodic Phrase Boundaries Revisited. *Tsinghua Science and Technology*, [s.l.], v. 13, n. 4, p. 500-509, 2008.

TSENG, C. Y.; FU, B. Duration, Intensity and Pause Predictions in Relation to Prosody Organization. In: EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY, INTERSPEECH, 9th, 2005. Lisboa. *Proceedings...* Lisboa: SISCA, 2005. p. 1405-1408. Available at: <<http://www.ling.sinica.edu.tw/eip/FILES/publish/2007.4.12.99500673.0143164.pdf>>. Retrieved on: Dec. 1, 2016.

TYLER, J. Prosodic Correlates of Discourse Boundaries and Hierarchy in Discourse Production. *Lingua*, [s.l.], v. 133, p. 101-126, 2013.

Van DONZEL, M. E. Prosodic Aspects of Information Structure in Discourse. Den Haag: Holland Academic Graphics/IFOTT, 1999.

WAGNER, A. Acoustic Cues for Automatic Determination of Phrasing. In: SPEECH PROSODY 2010 – INTERNATIONAL CONFERENCE, 5th., 2010, Chicago. *Proceedings...* Chicago: ISCA, 2010. Available at: <https://www.isca-speech.org/archive/sp2010/papers/sp10_196.pdf>. Retrieved on: Sept. 2018.

WAGNER, M.; WATSON, D. G. Experimental and Theoretical Advances in Prosody: A Review. *Language and Cognitive Processes*, [s.l.], v. 25, n. 7-9, p. 905-945, 2010.

WIGHTMAN, C. W.; SHATTUCK-HUFNAGEL, S.; OSTENDORF, M.; PRICE, P. J. Segmental Durations in the Vicinity of Prosodic Phrase Boundaries. *The Journal of the Acoustical Society of America* [s.l.], v. 91, n. 3, p. 1707-1717, 1992.

XU, Y. In Defense of Lab Speech. *Journal of Phonetics*, [s.l.], v. 38, n. 3, p. 329-336, 2010.

YANG, Y.; WANG, B. Acoustic Correlates of Hierarchical Prosodic Boundary in Mandarin. In: SPEECH PROSODY, 2002, Aix-en-Provence. *Proceedings...* Aix-en-Provence: Laboratoire Parole et Langage, 2002.

YOON, T-J.; CHAVARRÍA, S.; COLE, J.; HASEGAWA-JOHNSON, M. Intertranscriber Reliability of Prosodic Labeling on Telephone Conversation Using ToBI. In: SPEECH PROSODY INTERNATIONAL CONFERENCE, 2004. Nara, Japan. *Proceedings...* Nara: ISCA, 2004. p. 2722-2732.

ZATORRE, R. J. Cerebral Correlates of Human Auditory Processing: Perception of Speech and Musical Sounds. In: SYKA, J. (Ed.). *Acoustical Signal Processing in the Central Auditory System*. Prague: Plenum Press, 1997. p. 453-468.



FormantPro as a Tool for Speech Analysis and Segmentation

FormantPro como uma ferramenta para a análise e segmentação da fala

Yi Xu

Department of Speech, Hearing and Phonetic Sciences, University College London / UK
yi.xu@ucl.ac.uk

Hong Gao

English Department, Sichuan University, Chengdu / China
395478712@qq.com

Abstract: This paper introduces FormantPro, a Praat-based tool for large-scale, systematic analysis of formant movements, especially for experimental data. The program generates a rich set of output metrics, including continuous contours like time-normalized formant trajectories and formant velocity profiles suitable for direct graphical comparisons, and discrete measurements suitable for statistical analysis. It also allows users to generate mean trajectories and discrete measurements averaged across repetitions and speakers. As an illustration of its usage, data from a preliminary study of syllable segmentation in Mandarin were presented. The alignment of continuous formant trajectories enabled by FormantPro provides evidence that the temporal scopes of consonants and vowels are very different from those based on conventional views, and that acoustic and articulatory boundaries of segments are fundamentally similar.

Keywords: FormantPro; formant trajectories; syllable segmentation.

Resumo: Este artigo apresenta o FormantPro, uma ferramenta que roda no Praat, dedicada à análise sistemática e em larga escala dos movimentos de formantes, especialmente para dados de natureza experimental. O programa gera um rico conjunto de métricas de saída, incluindo contornos contínuos, como as trajetórias de formantes

normalizadas temporalmente e perfis de velocidade de formantes adequados para comparações gráficas diretas, bem como medidas discretas adequadas para a análise estatística. O programa também permite aos usuários gerar médias de trajetórias e medidas discretas calculadas a partir das médias de repetições e de falantes. Como ilustração da sua usabilidade, dados preliminares de um estudo sobre segmentação silábica em mandarim foram apresentados. O alinhamento de trajetórias contínuas de formantes geradas pelo FormantPro oferecem evidência de que os escopos temporais de consoantes e vogais são muito diferentes daqueles baseados em visões convencionais, e de que as fronteiras acústicas e articulatórias dos segmentos são fundamentalmente semelhantes.

Palavras-chave: FormantPro; trajetórias dos formantes; segmentação silábica

Submitted on January 8th, 2018

Accepted on June 7th, 2018

1. Introduction

Researchers frequently face a dilemma when it comes to taking formant measurements. Done too sparsely, important details may be missed; but continuous formant tracks are just too hard to process on a large scale. As a result, continuous formant contours are mostly used only as illustrations rather than as data in the literature. The benefit of analyzing fully continuous formants is evident from the classic work of Öhman (1966), whose insights on coarticulation gained from hand-traced formant trajectories are relevant even to the present day. But manual tracking of formants would no longer meet today's standards. Rapid technological advances have made automatically extracted continuous formants tracks easily available, yet they are still hard to use in systematic comparisons. The main difficulty is that when utterances differ in duration, it is hard to be sure whether we are comparing like with like as far as continuous trajectories are concerned.

FormantPro, available at <http://www.homepages.ucl.ac.uk/~uclyyix/FormantPro/>, is a software tool developed chiefly to address this dilemma. It is written as a Praat script (BOERSMA, 2001)—so that no programming is required of the users—for large-scale, systematic experimental studies of formant movements. FormantPro was first developed in 2007 (XU, 2007), and has been available online since 2013. The dedicated web page lists step-by-step instructions on how to use the script, how to read its output, as well as relevant information on time-normalization. The script has been used to generate results in a number of publications both by ourselves and by other researchers (e.g., BERKSON *et al.*, 2017; CHENG; XU, 2013; GAO; XU, 2013; LEE; MOK, 2017; LIU; LIANG, 2016; XU, 2007).

FormantPro applies time-normalization to extract the same number of evenly spaced formant values from each temporal interval, which allows users to treat any hypothetical unit as being temporally equivalent. The time-normalization algorithm was similar to that of ProsodyPro, a script for F_0 analysis (XU, 2013), except that the default number of normalized points is 20 instead of 10 due to faster segmental than F_0 changes in articulation (CHENG; XU, 2013; XU, 2007). The script further enables users to average the time-normalized formant as well as formant velocity trajectories across repetitions or even speakers. When plotted graphically, the trajectories can be compared between experimental conditions in a manner that is even more straightforward than in Öhman (1966), i.e., to overlay them in the same plot, as shown in the many examples presented later in this paper.

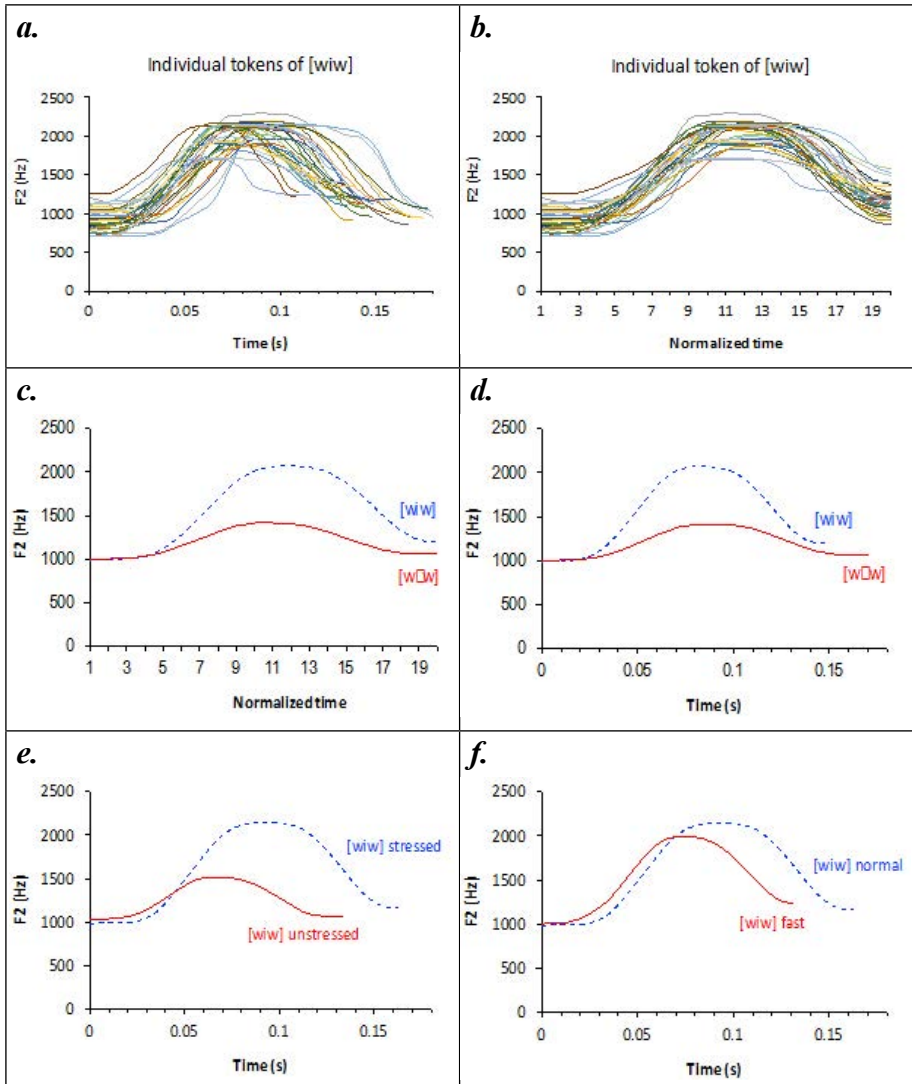
2. Dilemma and solution by FormantPro

Figure 1 shows F2 trajectories of [wiw] and [wow] in American English, generated by FormantPro, and plotted in three different time scales. The raw data are from utterances produced by 7 male speakers of American English. In Figure 1a, trajectories of 33 individual utterances of [wiw] are plotted in real time relative to the onset of the first F2 minimum. As can be seen, the trajectories vary extensively in duration. This makes it hard to see if there is any clear consistency across the individual tokens.

In Figure 1*b*, the same trajectories are time-normalized, i.e., consisting of 20 evenly spaced points. The consistency across the individual tokens now becomes much more apparent. The individual trajectories, however, even when time-normalized, are still not ideal for making cross-category comparisons. But given that they all consist of the same number of points, it is possible to average them at each and every point. The resulting mean trajectories can then be easily compared to each other when drawn in the same graph, as can be seen in Figure 1*c*, where the mean F2 trajectories of [wiw] and [waw] are plotted over normalized time.

A seeming disadvantage of time-normalization is that some of the original timing information may be lost. But this is not necessarily the case, as timing can also be abstracted. That is, like the formant values, the time value at each of the 20 points can be averaged across the individual tokens. In Figure 1*d*, the same [wiw] and [waw] trajectories are plotted over *averaged real time* across all individual tokens. Here the differences in terms of both curvature and timing of F2 movements between the two syllables can be easily seen. Likewise, the effects of stress and speech rate on the same syllable ([wiw]) can be easily seen when the F2 trajectories are plotted over mean averaged time in Figure 1*e* and 1*f*.

FIGURE 1 – F2 trajectories of [wiw] and [wɔw] produced by 7 male speakers of American English. In *a* and *b*, trajectories of all the individual utterances are plotted either in real time relative to the onset of the first F2 minimum (*a*) or in normalized time (*b*). In *d*, *e* and *f*, the *y* values of the trajectories are the mean F2 averaged at each of the 20 time-normalized points across all tokens by all speakers, but the *x* values are the mean times averaged also across all the repetitions and speakers at each of the 20 points. All contours are generated by FormantPro.



Another advantage of directly comparing continuous formant trajectories is that it allows one to clearly see where the largest differences are between the contrasting conditions, as the full time-course of the trajectories is immediately visible (Figure 1c-f). This enables one to make well-informed decisions when choosing measurements for statistical comparisons. Without such trajectory comparisons, decisions about where to take measurements are often made blindly, and the detection of critical differences is often a hit-and-miss game.

3. Usage and Features

FormantPro is written as a Praat script, which makes it executable on most of the major operating systems, including Mac, Windows and Unix. Written with large-scale systematic studies in mind, it maximizes efficiency of data processing by automating tasks that do not require human judgment, and by saving analysis output in formats that are ready for graphical and statistical analysis. More specifically, FormantPro allows users to:

- Manually segment and label intervals for each sound file, as illustrated in Figure 2,
- Cycle through all sound files in a folder without using menu commands, see Figure 3,
- Get maximum formant, minimum formant, mean formant, maximum formant velocity, duration and mean intensity from each labeled interval in each sound,
- Collect results from all individual sounds in a folder into a set of ensemble files that contain measurements of F1, F2, F3 and F2_3 in each interval of each sound file:
 1. meanformant.txt — mean values of the formants (Hz)
 2. maxformant.txt — maximum values of the formants (Hz)
 3. minformant.txt — minimum values of the formants (Hz)
 4. maxformantvelocity.txt — maximum velocity of the formants (Hz/s)
 5. formant.txt — time-normalized formants (Hz)

- 6. normtime_barkformant.txt — time-normalized formants in the Bark scale (Bark)
- 7. formantvelocity.txt — time-normalized formant velocity (Hz/s), and
- Get mean time-normalized formants, time-normalized formant velocities and actual times corresponding to the time-normalized formant and velocity points, averaged across repetitions as well as speakers.

FIGURE 2 – A TextGrid window with hand-labelled segmentation. FormantPro generates continuous as well as discrete measurements only for the labelled intervals. This allows users to obtain measurement generation only from regions of interest.

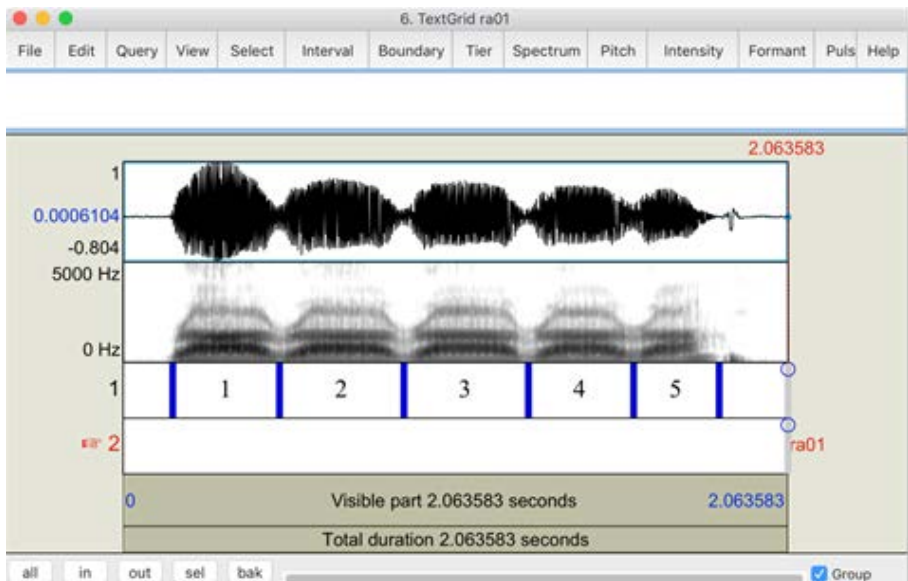
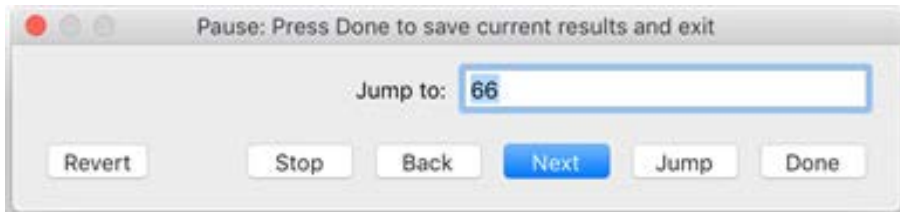
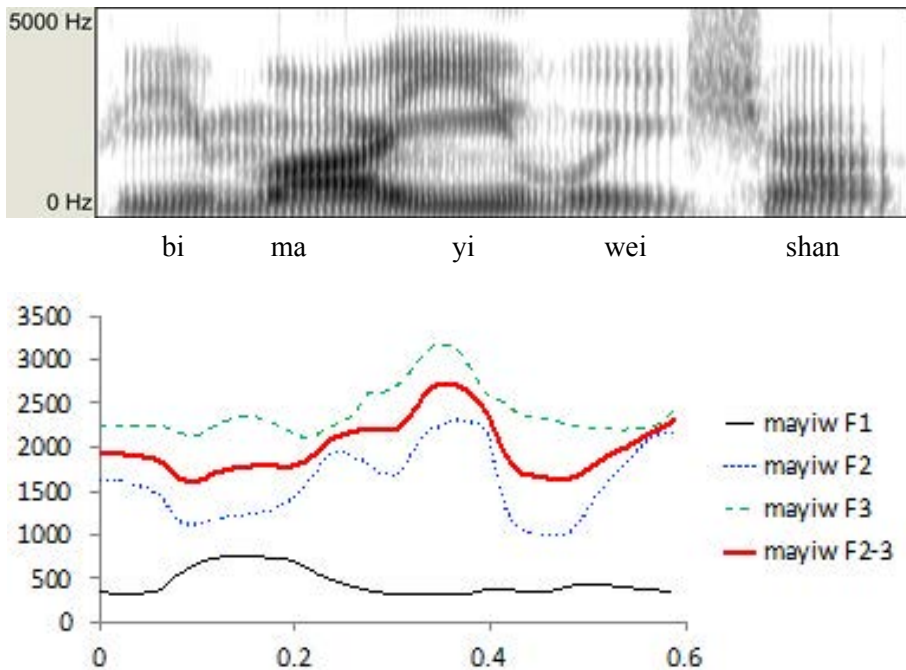


FIGURE 3 – The Pause window that controls the flow of the interactive segmentation and annotation. As the user moves forward, backward or jump to a particular sound file, the segmentation and measurements of the current file are automatically saved.



The continuous trajectories of F1, F2, F3 and F2_3 are generated with Praat's built-in "To Formant (burg)..." function. Here F2_3 = mean (F2, F3) is an unconventional one. It is motivated by the well-known problem of abrupt shifts of affiliation of formant with resonance cavity as vocal tract shape changes smoothly, e.g., between [i] and [a] (STEVENS, 1998). Averaging F2 and F3 can partially reduce the effects of the sudden shifts. Whether this measurement is advantageous over measuring F2 and F3 separately is an empirical matter. Data from one of our own studies (GAO; XU, 2013) seem to show partial support for this hypothesis. Making this measurement available in FormantPro will allow users to further test the hypothesis.

FIGURE 4 – Top: Spectrogram of Mandarin sentence “Bǐ Mǎyí wěishàn” [more hypocritical than Aunt Ma]. Bottom: Mean formant tracks of 10 repetitions by a male speaker of Mandarin.



Time-normalization, however, requires users to define the temporal domain of normalization. In FormantPro this is done by inserting interval boundaries in the TextGrid of an utterance. Technically FormantPro allows user to freely annotate the temporal domains of normalization, e.g., segment, syllable or even word. But meaningful time-normalization can be obtained only if there are good reasons to believe that the formant trajectories in the unit are consistently produced, which is both a theoretical and empirical matter.

To segment continuous speech into discrete units, one of the critical questions is, what is the acoustic correlate of a phonetic unit? In the current practice, the answer is that a unit, such as a consonant or vowel, is what is delimited by the landmarks (STEVENS, 2002) on a spectrogram, such as abrupt spectral shift, onset and offset of oral closure, etc. (TURK; NAKAI; SUGAHARA, 2006), which is also what sounds like that phone when isolated from the acoustic stream (ZUE *et*

al., 1990). For example, in Figure 4, the [i] in “bǐ” is to be delimited by the first and second abrupt spectral shifts after the consonant release and before the nasal murmur; the [m] in “má” is delimited by the onset and offset of the nasal murmur; and the [ʃ] is delimited by the onset and offset of the frication. This segmentation scheme, however, leaves many cases unresolved. In Figure 4, for example, the exact offset of [a], the onset as well as the offset of [ji], and the onset of [wei] is by no means clear. The vagueness of their segmentation has led to explicit advice to avoid the glides when precise duration measurements are needed (TURK; NAKAI; SUGAHARA, 2006).

From an articulatory perspective (SALTZMAN; MUNHALL, 1989; XU; WANG, 2001), however, unit boundaries can be defined rather differently. That is, the onset of a unit should be the moment when the articulators start to move toward their target positions defined by its canonical form, and the offset of the unit should be the moment when the articulators start to move away from those positions. The canonical form of a monophthong vowel would be the ideal vocal tract shape that generates the steady-state prototypical formant pattern, and the canonical form of a consonant would be the ideal closure or constriction at the appropriate place of articulation. The movements toward these targets take time, and *it is the time course of the movement that should be considered as the interval of the unit* (XU; LIU, 2007). In other words, a unit is delimited by the onset and offset of the movement toward its target.

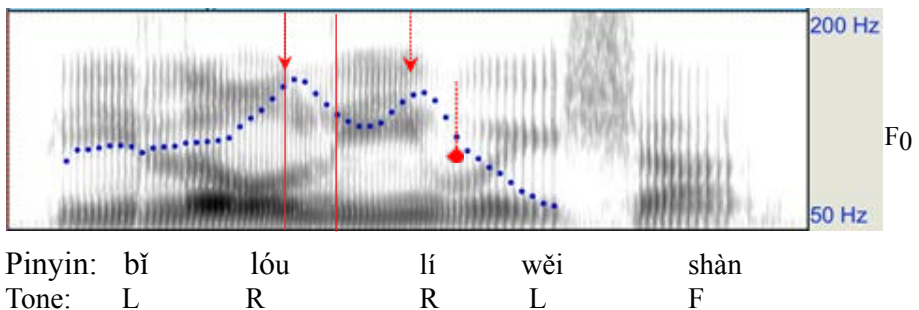
It is not always easy to identify the onset and offset of a movement, however. In the following, we discuss a method that uses a combination of graphical comparison of formant trajectories and F_0 -segment alignment to determine the temporal scope of segments in Mandarin. The first component of the method is minimal contrast comparison of continuous trajectories, which has been applied extensively on F_0 analysis for tone and intonation (e.g., XU, 1999; XU; XU, 2005). For segmental analysis, minimal contrast comparison has been applied in analysis of articulatory data (BOYCE; KRAKOW; BELL-BERTI, 1991; GELFER; BELL-BERTI; HARRIS, 1989), but it has not been widely used in formant analysis, partly because of a lack of convenient tools, which is no longer the case with the availability of FormantPro. The key to minimal contrast comparison of trajectories is to graphically compare the contrasting movements in question in identical or near-identical contexts. This way, aspects of the trajectories that are due to contextual variations are made

identical, so that the differences between the contrasting trajectories become unambiguous.

The second component of the method is to use F_0 events, such as turning points, as temporal anchor points to align the contrasting trajectories. The rationale comes from findings of consistent F_0 -segment alignment in various languages (ARVANITI *et al.*, 1998; LADD *et al.*, 1999; SCHEPMAN *et al.*, 2006; XU, 1998). That is, other things being equal, certain F_0 turning points regularly occur near the onset or offset of a syllable. In Mandarin, for example, the F_0 of the Rising tone (T2) consistently peaks right after syllable offset when followed by a Low or Rising tone. In Figure 5, for example, where the second and third syllable both have the Rising tone, the first F_0 peak occurs right after the onset of the [l] murmur in “lí”.

The significance of the constant F_0 -segment alignment is that it goes both ways. That is, it is also the case that the segmental events involved are likewise aligned to the F_0 events. This further means that F_0 events can be used to determine segmental alignment when there is a lack of landmarks, e.g., in the case of glides and approximants. For example, as found in Xu and Liu (2007), when the F_0 peak is used as the temporal reference, the equivalent of the [l] closure onset in “wěi” would be at the second arrow in Figure 5, as opposed to the low turning point of F2 at the diamond head arrow which has been suggested as a landmark (STEVENS, 2002).

FIGURE 5 – Spectrogram of “Bǐ Lóulí wěishàn” [more hypocritical than Louli], with pitch track (blue speckles) generated by Praat. The two vertical lines mark the onset and offset of [l] closure.



3.1 An illustrative experiment

A preliminary experiment was designed to assess the temporal scope of consonants and vowels in CV syllables in Mandarin. One set of the stimuli is shown in Table 1. The stimuli are $C_1V_1\#C_2V_2$ disyllabic words that form four triplets, each shown in a row in the table. In each triplet, the first two words differ from each other in C_2 : [j] vs. [l], while the second two differ in V_2 : [i] vs [u]. The first two words therefore form a minimal pair for which the divergent point of their F2 trajectories would indicate the onset of C_2 , and the second two words form a minimal pair for which the divergent point of F2 would indicate the onset of V_2 . The two consonants are both sonorants that do not involve full closure of the oral cavity, thus allowing continuous formant movements to be seen during the consonantal constrictions. In addition, all the words have the Rising tone (Tone 2, with the tone mark [ˊ]) on both syllables, so as to allow the occurrence of two F_0 peaks that can serve as time references for the onset and offset of the second syllable.

TABLE 1 – Disyllabic words in 3-way contrasts: [j]/[l] as initial C between words in the first two columns, and [i]/[u] as nuclear V between words in the last two columns.

Pinyin	Chinese	Pinyin	Chinese	Pinyin	Chinese
léiyí	雷姨	léilí	雷黎	léilú	雷庐
máyí	麻姨	máilí	麻黎	máilú	麻庐
lóuyí	娄姨	lóulí	娄黎	lóulú	娄庐
lúyí	卢姨	lúlí	卢黎	lúlú	卢庐

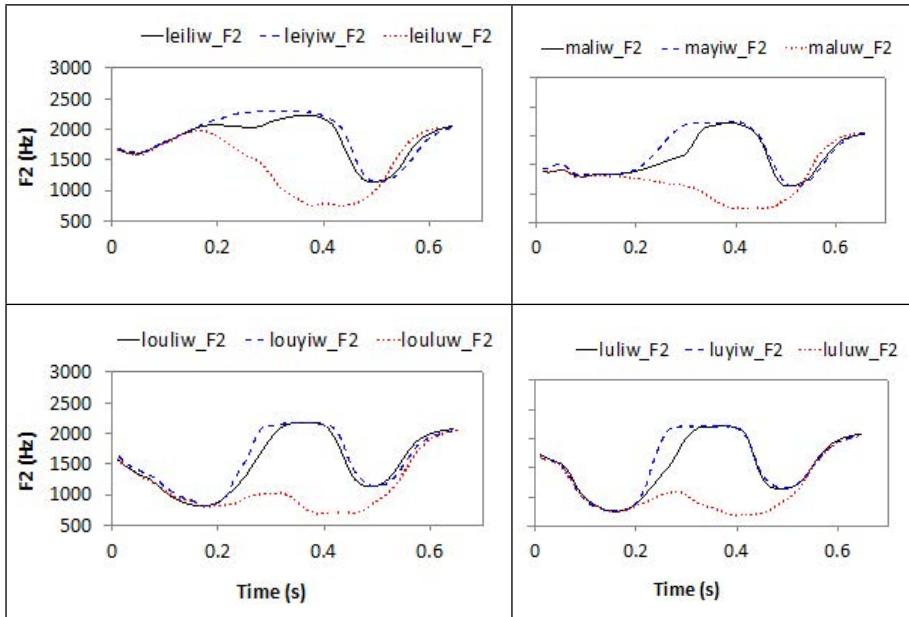
Three male native speakers of Mandarin read aloud the triplets, each in the carrier “Bǐ ___ wěishàn” [more hypocritical than ___], with 8 repetitions each, in separate randomized blocks. Their formant trajectories were extracted with FormantPro, and their F_0 patterns with ProsodyPro (XU, 2013). A separate Praat script was written to align the formant trajectories with respect to the F_0 peaks associated with the two Rising tones in each word. All the formant trajectories were taken at 20 evenly spaced locations in each syllable after the F_0 -based boundary adjustment. Mean trajectories were then obtained by averaging across the repetitions as well as speakers. At the same time, time values at each of the 20 points

were also averaged across the repetitions and speakers, which will serve as time axes for some of the formant plots in the analysis.

3.2 Graphical analysis and discussion

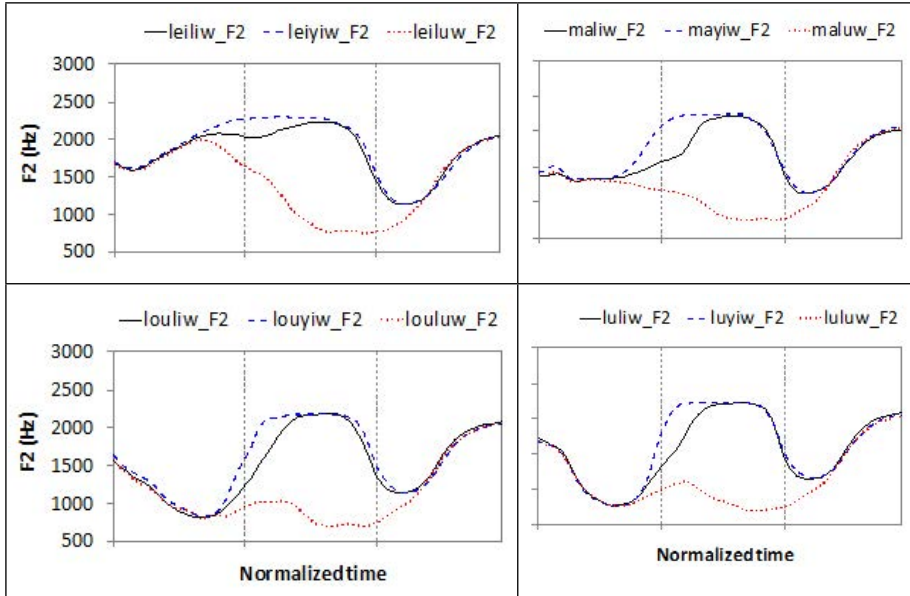
Figure 6 displays grand mean F2 trajectories of the four triplets in Table 1. In each plot, the solid and dashed lines differ in the initial consonants: [l] vs. [j], and the point at which the two trajectories start to diverge would indicate the onset of both consonants, as it is where the articulatory movements start to move toward their respective targets. The solid and dotted lines, on the other hand, differ in the vowels of the second syllable: [i] vs. [u], and the point at which the two trajectories start to diverge would indicate the onset of both vowels. Strikingly, in each case the vowel divergent point occurs at about the same time as the consonant divergent point. Since the contrasting syllables are [li] and [lu], the V approaching movements actually also includes movements toward[l], as revealed by the contrast between [li] and [ji]. In other words, contrary to the conventional view that the acoustic onset of the vowel starts much later—i.e., at the voice onset—than that of the consonant in a CV syllable, the F2 dynamics suggests that the two may actually start at the same time.

FIGURE 6 – Mean F2 trajectories of four triplets in Table 1, plotted on mean time relative to the onset of [l] or [m] in the first syllable of the target word. Both F2 and time are averaged across 8 repetitions by 3 male speakers.



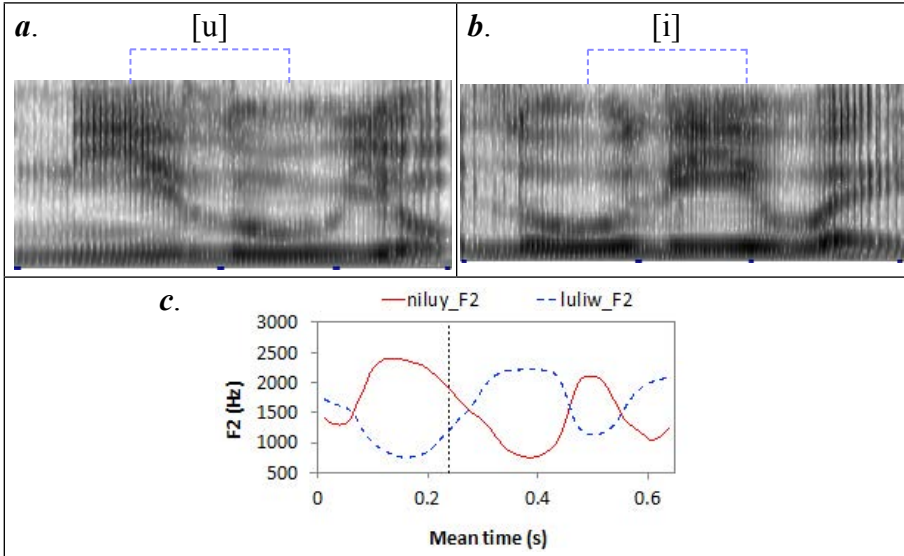
To further explore the exact location of the common starting point of C and V, the F_0 -aligned F2 trajectories are plotted on normalized time in Figure 7. The two vertical lines in each plot are at the F_0 peaks, which divide the formant trajectories into three intervals, each corresponding to one of the conventional syllables. The time-normalized F2 trajectories show greater consistencies within all the triplets than those on averaged real time in Figure 6, indicating the relevance of syllable as a unit of articulatory target approximation. Most relevantly, the joint onset of C2 and V2 movement toward their respective targets can now be seen as well before (about 50–100 ms based on a preliminary estimate) the conventional syllable onset.

FIGURE 7 – Mean time-normalized F2 trajectories of the four triplets in Table 1, averaged across 8 repetitions by 3 male speakers. The two vertical lines in each plot are at the F₀ peaks.



The significance of the new estimate of vowel onset in CV syllables is even more striking when formant movements are spectrally visible across conventional boundaries thanks to the articulatory transparency of [l], as can be seen in Figure 8. In (a), F2 moves continuously from its highest position in the middle of the vocalic section of [ni] to the middle of the vocalic section of [lu]. Assuming that movement toward a target is the scope of a vowel as hypothesized, this entire downward movement would constitute the temporal scope of [u]. Likewise, the entire rising movement of F2 in Figure 8b would constitute the temporal scope of [i]. These scopes are strikingly different from the conventional segmentation as marked by the transcriptions below the spectrograms. When the entire formant trajectories of [li] and lu] are laid on top of each other in Figure 8c, they form a mirror image that makes the temporal domains of the vowels even less ambiguous.

FIGURE 8 – (a, b) Spectrograms of [ni lu jou] and [lu li wei] in Mandarin, with conventional segmentation at the bottom and formant-dynamics-based segmentation of [u] and [i]. (c) Mean F2 trajectories of the two words averaged across 8 repetitions by 3 male speakers, plotted on mean time of all tokens of the two words.



What we have demonstrated above is not entirely new, because Öhman (1966) already reported that articulatory movements toward the nuclear vowel in a CV syllable may start during the intervocalic consonant. Research based on articulatory phonology (BROWMAN; GOLDSTEIN, 1992) and the task dynamic model (SALZMAN; MUNHALL, 1989) has also shown heavy overlap of C and V at the syllable initial position. However, in the widespread common practice, vowels are still routinely assumed to start at the consonant release, and any acoustic properties that may reflect the vowel during or before the initial consonant are attributed to anticipatory coarticulation (FOWLER; SALTZMAN, 1993; LINDBLOM; SUSSMAN, 2012). There may be two reasons for the endurance of the conventional segmentation. Firstly, the landmarks are just too visually compelling to ignore: How can the [u] in Figure 8a, for example, start from where the formants clearly indicate the vowel [i], and by so doing cross the entirety of the consonant [l] right in the middle? Secondly, the articulatory-based segmentation is often auditorily implausible: The [u] and [i] segments as suggested in Figure

7, for example, would both sound like two syllables due to the [l] closure in the middle. How can they be considered as corresponding to single vowels? For an articulatory-based acoustic segmentation to be sufficiently compelling, it is necessary to demonstrate that articulatory dynamics is in fact directly reflected in the acoustics. The direct visualization of continuous formant trajectories like those shown in Figures 6-8 generated by FormantPro allows us to see that articulatory dynamics is actually much more acoustically transparent than is generally believed. Thus there may be sufficient ground to assume articulatory and acoustic segmentations as fundamentally the same.

3.3 Caveats

The preliminary data in this section are presented mainly for illustrating the use of FormantPro. The methodology described is designed only for the specific question addressed in the study. In particular, two clarifications are in order. First, the use of F_0 as a reference is only a useful strategy rather than a mandatory requirement for formant analysis. What has been demonstrated is that, like for F_0 , the dynamic aspect of segmental articulation can be studied by examining continuous formant trajectories with the availability of FormantPro, and it is possible to also combine it with ProsodyPro to explore some questions in ways that go beyond what can be done with conventional methods.

Secondly, despite the preliminary evidence for simultaneous onset of consonant and vowels that is much earlier than those based on standard practice, it is not yet clear how the finding, if further confirmed, can be used in phonetic segmentation of speech utterances for annotation purposes. One possibility is to establish, through large-scale empirical testing, segmentation rules that can be easily applied in practice. For example, for simple CV syllables where the C is an obstruent consonant, the C-V co-onset point can be set at a fixed amount of time, e.g., 50 ms (which could be speaker-specific due to individual differences in articulation rate) ahead of the easily observable closure onset.

4. Conclusions

In this paper we have introduced FormantPro, a Praat-based research tool for systematic analysis of formants. The tool facilitates analysis of articulatory dynamics through direct comparison of

continuous formant trajectories. This is achieved by, among other things, allowing users to obtain time-normalized formant trajectories that can be averaged across repetitions as well as speakers. It also facilitates systematic analysis of large amount of experimental data by automating procedures that do not require human judgment and saving a variety of formant, duration and intensity measurements in formats that are ready or near-ready for statistical analysis. As an illustration, we also presented preliminary data from a study aimed at assessing the temporal scope of consonants and vowels in CV syllables in Mandarin. These data provide evidence that the temporal scope of vowel is much larger than what is mostly assumed in common practice, as it starts roughly at the same time as the initial consonant. The new evidence for the co-onset of C and V may lead to a new discussion of coarticulation that treats the identification of temporal scope of phonetic units as a prerequisite.

Authors' contribution

Y.X. developed FormantPro and conceived the experiment. H.G. carried out the experiment. Y.X. and H.G. jointly performed data analysis and co-wrote the paper.

References

- ARVANITI, A.; LADD, D. R.; MENNEN, I. Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics*, Elsevier, v. 36, p. 3-25, 1998.
- BERKSON, K.; DAVIS, S.; STRICKLER, A. What does incipient/ay/-raising look like?: A response to Josef Fruehwald. *Language*, Washington, v. 93, n. 3, p. e181-e191, 2017.
- BOERSMA, P. Praat, a system for doing phonetics by computer. *Glott International*, Blackwell Publishing, v. 5, n. 9/10, p. 341-345, 2001.
- BOYCE, S. E.; KRAKOW, R. A.; BELL-BERTI, F. Phonological under specification and speech motor organization. *Phonology*, Elsevier, v. 8, p. 210-236, 1991.
- BROWMAN, C. P.; GOLDSTEIN, L. Articulatory phonology: An overview. *Phonetica*, International Society of Phonetic Sciences, v. 49, p. 155-180, 1992. Doi: [10.1159/000261913](https://doi.org/10.1159/000261913)

CHENG, C.; XU, Y. Articulatory limit and extreme segmental reduction in Taiwan Mandarin. *Journal of the Acoustical Society of America*, v. 134, n. 6, p. 4481—4495, 2013.

FOWLER, C. A.; SALTZMAN, E. Coordination and coarticulation in speech production. *Language and Speech*, Sage Journals, v. 36, n. 2-3, p. 171-195, 1993.

GAO, H.; XU, Y. Coarticulation as an epiphenomenon of syllable-synchronized target approximation—Evidence from F0-aligned formant trajectories in Mandarin. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 135, Pt. 2, 2013.

GELFER, C. E.; BELL-BERTI, F.; HARRIS, K. S. Determining the extent of coarticulation: effects of experimental design. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 86, n. 6, p. 2443-2445, 1989.

LADD, D. R.; FAULKNER, D.; FAULKNER, H.; SCHEPMAN, A. Constant “segmental anchoring” of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 106, p. 1543-1554, 1999.

LEE, A.; MOK, P. Acquisition of Japanese quantity contrasts by L1 Cantonese speakers. *Second Language Research*, Hong Kong, 2017. Doi: <http://dx.doi.org/10.1177/0267658317739056>

LINDBLOM, B.; SUSSMAN, H. M. Dissecting coarticulation: How locus equations happen. *Journal of Phonetics*, Elsevier, v. 40, n. 1, p. 1-19, 2012.

LIU, H.; LIANG, J. Vowels as acoustic cues for sub-dialect identification in Chinese. In: INTERNATIONAL SYMPOSIUM CHINESE SPOKEN LANGUAGE PROCESSING (ISCSLP), 10th., Tianjin, China, 2016. *Proceedings...* Tianjin, China: IEEE, 2016. p. 1-5.

ÖHMAN, S. E. G. Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 39, p. 151-168, 1966.

SALTZMAN, E. L.; MUNHALL, K. G. A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, Francis & Taylor Online, v. 1, p. 333-382, 1989.

SCHEPMAN, A.; LICKLEY, R.; LADD, D. R. Effects of vowel length and “right context” on the alignment of Dutch nuclear accents. *Journal of Phonetics*, Elsevier, v. 34, p. 1-28, 2006.

STEVENS, K. N. *Acoustic Phonetics*. Cambridge, MA: The MIT Press, 1998.

STEVENS, K. N. Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 111, p. 1872-1891, 2002.

TURK, A.; NAKAI, S.; SUGAHARA, M. Acoustic Segment Durations in Prosodic Research: A Practical Guide. In: SUDHOFF, S.; LENERTOVA, D.; MEYER, R. *et al. Methods in Empirical Prosody Research*. Berlin; New York: De Gruyter, 2006. p. 1-28.

XU, Y. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica*, Bankstown, Australia, v. 55, p. 179-203, 1998.

XU, Y. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, Elsevier, v. 27, p. 55-105, 1999.

XU, Y. ProsodyPro — A tool for large-scale systematic prosody analysis. In: TOOLS AND RESOURCES FOR THE ANALYSIS OF SPEECH PROSODY (TRASP 2013), Aix-en-Provence, France, 2013. *Proceedings...* Aix-en-Provence: [s.n.], 2013. p. 7-10.

XU, Y. How often is maximum speed of articulation approached in speech? *Journal of the Acoustical Society of America*, Acoustical Society of America, v. 121, Pt. 2, p. 3199-3140, 2007.

XU, Y.; LIU, F. Determining the temporal interval of segments with the help of F0 contours. *Journal of Phonetics*, Acoustical Society of America, v. 35, p. 398-420, 2007.

XU, Y.; WANG, Q. E. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, Elsevier, v. 33, p. 319-337, 2001.

XU, Y.; XU, C. X. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, Elsevier, v. 33, p. 159-197, 2005.

ZUE, V.; SENEFT, S.; GLASS, J. Speech database development at MIT: TIMIT and beyond. *Speech Communication*, v. 9, n. 3, p. 1-356, 1990.



Acoustic Models for the Automatic Identification of Prosodic Boundaries in Spontaneous Speech

Modelos acústicos para a identificação automática de fronteiras prosódicas na fala espontânea

Bárbara Helohá Falcão Teixeira

Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais / Brazil
barbaraheloha@gmail.com

Maryualê Malvessi Mittmann

Centro Universitário FACVEST, Lages, Santa Catarina / Brazil
Universidade do Vale do Itajaí, Itajaí, Santa Catarina / Brazil
mittmann@univali.br

Abstract: This work presents the results of the analysis of multiple acoustic parameters for the construction of a model for the automatic segmentation of speech in tone units. Based on literature review, we defined sets of acoustic parameters related to the signalization of terminal and non-terminal boundaries. For each parameter, we extracted a series of measurements: 6 for speech rate and rhythm; 34 for duration; 65 for fundamental frequency; 4 for intensity and 2 measurements related to pause. These parameters were extracted from spontaneous speech fragments that were previously segmented into tone units, manually performed by 14 human annotators. We used two methods of statistical classification, Random Forest (RF) and Linear Discriminant Analysis (LDA), to generate models for the identification of prosodic boundaries. After several phases of training and testing, both methods were relatively successful in identifying terminal and non-terminal boundaries. The LDA method presented a higher accuracy in the prediction of terminal and non-terminal boundaries than the RF method, therefore the model obtained with LDA was further refined. As a result, the terminal boundary model is based on 20 acoustic measurements and shows a convergence of 80% in relation to boundaries identified by annotators in the speech sample. For non-terminal boundaries, we arrived at three models that, combined, presented a convergence of 98% in relation to the boundaries identified by annotators in the sample.

Keywords: speech segmentation; prosodic boundaries; spontaneous speech.

Resumo: Este trabalho apresenta os resultados da análise de múltiplos parâmetros acústicos para a construção de um modelo para a segmentação automática da fala em unidades tonais. A partir da investigação da literatura, definimos conjuntos de parâmetros acústicos relacionados à identificação de fronteiras terminais e não terminais. Para cada parâmetro, uma série de medidas foram extraídas: 6 medidas de taxa de elocução e ritmo; 34 de duração; 65 de frequência fundamental; 4 de intensidade e 2 medidas relativas às pausas. Tais parâmetros foram extraídos de fragmentos de fala espontânea previamente segmentada em unidades tonais de forma manual por 14 anotadores humanos. Utilizamos dois métodos de classificação estatística, *Random Forest* (RF) e *Linear Discriminant Analysis* (LDA), para gerar modelos de identificação de fronteiras prosódicas. Após diversas fases de treinamentos e testes, ambos os métodos apresentaram sucesso relativo na identificação de fronteiras terminais e não-terminais. O método LDA apresentou maior índice de acerto na previsão de fronteiras terminais e não-terminais do que o RF, portanto, o modelo obtido com este método foi refinado. Como resultado, O modelo para as fronteiras terminais baseia-se em 20 medidas acústicas e apresenta uma convergência de 80% em relação às fronteiras identificadas pelos anotadores na amostra de fala. Para as fronteiras não terminais, chegamos a três modelos que, combinados, apresentaram uma convergência de 98% em relação às fronteiras identificadas pelos anotadores na amostra.

Palavras-chave: segmentação da fala; fronteiras prosódicas; fala espontânea.

Submitted on January 12th, 2018

Accepted on June 17th, 2018

1 Introduction

This paper presents results from an investigation that aims at the construction of a model for spontaneous speech segmentation based on acoustic parameters. Natural speech is segmented into intonation units, delimited by prosodic boundaries that signal the conclusion or continuity of discourse. These boundaries are acoustically signaled by parameters such as pitch reset, pauses and syllabic lengthening, among others.

Although we have by now a good overall understanding of different parameters involved in speech segmentation (for a review, see MITTMANN; BARBOSA, 2016), there is no approach that allows us to integrate them into a model that could be applied for the automatic detection of prosodic boundaries in spoken texts. Moreover, discrimination between terminal (conclusive) and non-terminal boundaries is essential, since this information

is key to the correct identification of syntactic relations inside the utterance, as well as its pragmatic meaning (for a discussion and demonstration regarding this argument see MONEGLIA, 2011; RASO; VIEIRA, 2016).

Therefore, our research aims to develop a tool that aggregates acoustic data of multiple acoustic parameters together with information about boundary type (terminal or non-terminal) obtained from human annotation of spontaneous speech input. The results will allow the creation of a computational tool for the automatic (or, at least, semiautomatic) detection of prosodic boundaries. Such tool would aid the compilation of spontaneous speech corpora, since it can make the speech segmentation process faster, saving time and effort, what could contribute to corpus linguistics in general.

This research represents an advance not only in the technological aspects of speech processing, but it implies in a better understanding about speech segmentation phenomena. Thus, we hope to contribute to the theory of speech, by promoting more accurate descriptions of phonetic phenomena involved in the linguistic processes that guide production and perception of terminal and non-terminal prosodic boundaries in spontaneous speech.

Prosodic segmentation of speech implies a series of methodological challenges. Boundaries are always signaled by phonetic phenomena, but those vary substantially in spontaneous speech. Working with non-natural and manipulated data provides comparable, high acoustic quality data, but represents enormous limitations when compared with the phenomena that occur in spontaneous, natural occurring data.

When we choose to work with spontaneous speech data, finding comparable speech segments is very difficult, and data with high acoustic quality may be hard to obtain. Besides, controlling variables one by one is not a possibility with spontaneous speech data. For these reasons, we employed statistic classification methods to arrive at models for automatic identification of terminal and non-terminal boundaries in spontaneous speech.

2. Speech segmentation based on prosodic cues

Speech is usually described as a “flow”, and identifying its segmental units is not a simple, straightforward, task. Segmentation of speech has been studied according to different theoretical perspectives. The syntactic approach proposes that the syntactic level of the sentence corresponds to a phonological level of the intonational phrase (COOPER; PACCIA-COOPER, 1980; SELKIRK, 2005). The pragmatic perspective states that prosodic parsing organizes speech by the demarcation of discourse or

information units (CRESTI, 2000; HALLIDAY, 1965; SZCZEPEK REED, 2012). The cognitive view studies the relation among units of speech and units of language processing by the brain (BYBEE, 2010; CHAFE, 1994; CROFT, 1995). Finally, the conversation analysis approach claims that breaks in the speech flow – cesuras – are granular by nature and the units they encompass cannot be discriminated into atomized categories, and so, segmentation analysis should regard the boundaries themselves instead of the units (AUER, 2010; BARTH-WEINGARTEN, 2016).

In this paper, we propose that a model for speech segmentation should primarily identify prosodic boundaries that listeners recognize in spontaneous speech. Perception of prosodic boundaries may vary, since there are boundaries that are more clearly signaled, or more prominent, than others.

Corpus-based observations and experimental research (BARBOSA, 2008; COUPER-KUHLEN, 2006; FUCHS; KRIVOKAPIC; JANNEDY, 2010; MITTMANN et al., 2010; MO, 2008; MONEGLIA; CRESTI, 2006; SCHUETZE-COBURN; SHAPLEY; WEBER, 1991; SWERTS; COLLIER; TERKEN, 1994) allow us to distinguish two boundary macrotypes: boundaries that signal discourse closure and boundaries that are not correlated to a closure. The first type is referred to in this paper as terminal boundary, and the second, non-terminal boundary. This two boundary macrotypes will be further discussed in the following sections. We also assume that the units delimited by those boundaries are the key for speech interpretation, as they mostly correspond to the organization of speech into information units (CRESTI; MONEGLIA, 2010; MONEGLIA, 2006), inside of which the morphosyntactic relations occur.

Most models for automatic speech segmentation aim to identify boundaries between phones and words, and then bootstrap syntactic relations from word sequences to arrive to the uttered sentence. The acoustic speech signal contains much of the information needed for extraction of the phonetic structure of the linguistic message (FOWLER, 1984). However, speech sounds blend together and cannot easily be separated, not only within words but also across words, due to speech coarticulation. Lexical, syntactic, and acoustic information are usually cues employed for word recognition, but some of them may work only for certain languages and all of them may be misleading in normal speech (for a discussion, see SANDERS; NEVILLE, 2000). Also, in spontaneous speech, syntactic and semantic relations can only be properly interpreted within the scope of units defined by prosody, such as utterances and tone units (BOSSAGLIA, 2016; CRESTI, 2014; IZRE'EL, 2011; MONEGLIA, 2011; RASO; VIEIRA,

2016). For these reasons, automatic models for speech segmentation that use the word as the base for segmentation are very complex and do not seem to be a good solution to spontaneous speech analysis. The best starting point for segmentation of the speech signal is prosody.

The role of prosody in speech segmentation is well acknowledged in linguistics literature. Among the functions of prosody, we can distinguish demarcation, i.e., marking boundaries of prosodic constituents, such as syllables, phonological words and groupings of speech in tone units (BARBOSA, 2012). According to Cruttenden (1997), a set of internal and external criteria can be applied to prosodic boundary identification. Among external criteria there are pre-boundary syllabic lengthening, presence of silent pause, changes in pitch level or direction. An example of internal criterion is the presence of a prominent syllable, called a nucleus, with a pitch movement. Crystal (1969) argues that aspiration is also a possible relevant acoustic parameter for boundary marking.

Considering the difficulty of applying these criteria in spontaneous speech, Cruttenden (1997) recommends the adoption of grammatical criteria, arguing that prosodic boundaries often co-occur with syntactic constituent limits. However, spontaneous speech corpora data show that, in many cases, there is no co-occurrence between prosodic and syntactic boundaries of constituents. Besides, adoption of grammatical criteria for prosodic boundary identification should be avoided, because it implies in describing a phonetic phenomenon by means of morphosyntactic categories.

Prosodic boundaries can be more or less perceptually prominent. The fact that boundaries do not constitute a categorical perceptual entity (AUER, 2010; BARTH-WEINGARTEN, 2016; BIRKNER, 2006; BOLINGER, 1972) is one of the reasons why their study is so complex. If some prosodic boundaries are very prominent and perceived by almost everyone, others show much less perceptual agreement among different speakers/listeners. When that is the case, many scholars end up making decisions about boundary marking based on linguistic theory, thus creating a circularity effect, as discussed by Brown *et al.* (1980) and Peters, Kohler and Wesener (2005). Therefore, in agreement with Auer and Barth-Weingarten, we believe that it is important to study the acoustic features that signal prosodic boundaries independently of the analysis of the units delimited by them.

According to Du Bois *et al.* (1992), prototypical prosodic units present: a coherent and unified pitch contour, pitch reset to the base level at the beginning of the unit, pause at the beginning of the unit, a high speech rate at the initial syllables of the unit, lengthening of one

or more syllables on the final portion of the unit. However, prosodic boundaries usually do not present all these features, so it is possible to divide them into two boundary types: “Full” boundaries, which have all the prototypical characteristics, and “partial” boundaries, which present only some of the prototypical characteristics. Because of the less precise demarcation of some boundaries, Du Bois (2008) complements the list of acoustic cues, including boundary tone, number of pitch accents, creaky voice, turn taking, rhythm and pitch changes.

This list of acoustic parameters related to boundary marking is supported by a great number of experimental research on various languages, such as English (COLE; SHATTUCK-HUFNAGEL; MO, 2010; MO; COLE; LEE, 2008)⁵⁴ excerpts, 11-55-s duration each, German (BATLINER *et al.*, 1995; FUCHS *et al.*, 2010; KOHLER; PETERS; WESENER, 2001), Dutch (BLAAUW, 1994; SWERTS, 1997; SWERTS; COLLIER; TERKEN, 1994), Portuguese (BARBOSA, 2008; RASO; MITTMANN; MENDES, 2015) showing the interdependence between f_0 and syllable-sized duration contours, showing the separate contributions of duration and f_0 at minor prosodic boundaries, presenting a semi-automatic method for analysing the correlation between f_0 and normalised syllablesized duration contours. Contrary to the observations in lab speech for isolated utterances, pitch accents are relatively frequent in BP (from 54 to 73 % of the phonological words and Mandarin (FON; JOHNSON; CHEN, 2011; TSENG; CHANG, 2008; TSENG *et al.*, 2005) syllable duration, pause duration, and syllable onset intervals (SOIs, just to cite a few. This variety of parameters shows how complex the acoustic correlates of boundaries are. Also, even if certain parameters had been shown to be strong correlates of boundaries, there is still no consensus regarding how much each individual parameter contributes to explain boundary perception. This occurs because, in many cases, a given parameter may be a very strong boundary predictor, but it could be completely absent in many other boundary positions. This problem is discussed in more detail by Mittmann and Barbosa (2016).

Another issue that adds up to this complexity regards to the type of boundary and how acoustic parameters correlate with each type. From a perceptual point of view, it seems evident that prosodic boundaries are not all of the same type. Researchers usually refer to boundaries associated with the perception of discourse completion or boundaries that signal discourse continuation (PIERREHUMBERT, 1980; PIKE, 1945; SZCZEPEK REED, 2004). Therefore, one would expect two

sets of acoustic correlates: one for conclusive boundaries, another for continuative boundaries. However, as we will discuss in the next sections, boundary typology is more complex than the conclusive-continuative dichotomy, and as our results show, it is not possible to arrive at two well-delimited groupings of prosodic parameters.

2.1 Terminal boundaries

The first macrotype of prosodic boundary refers to the ruptures in the speech flow that correspond to the perception of discourse closure or conclusion. These terminal boundaries signal the completion (in most cases) of an utterance, that is a linguistic entity that has prosodic and pragmatic autonomy in spoken discourse, as it expresses the completion of a speech act (AUSTIN, 1962; CRESTI; MONEGLIA, 2010). Some researchers refer to these units as “spoken sentences”, or “sentence-like units”, since, from the syntactic point of view, utterances not always correspond to the grammatical notion of “sentence”. Terminal boundaries delimit utterances, which may be (or may be not) further parsed into smaller units by means of non-terminal boundaries.

Example 1¹ illustrates an utterance delimited by what can be considered a prototypical terminal boundary. In our research, in this example, the boundary at the end of the utterance was identified as terminal by 14 out of 14 annotators (indicated by the red arrow in Figure 1). Figure 1 shows the soundwave, spectrogram, pitch contour and textgrid of example 1. Textgrid has five tiers, representing, from top to bottom:

- 1st– V-V tier: vowel to vowel² intervals with broad phonetic transcription in ASCII characters;
- 2nd – NTB tier: points indicate phonological words’ boundaries, numbers at each point indicate the number of annotators that perceived the point as a non-terminal boundary;
- 3rd – TB tier: points indicate phonological words’ boundaries, numbers at each point indicate the number of annotators that perceived the point as a terminal boundary;

¹ All examples come from the samples prepared for this research, based spontaneous speech corpus C-ORAL-BRASIL, as described in the “Methods” section.

² For clarification, see the “Methods” section.

4th – Pause tier: silent pauses intervals.

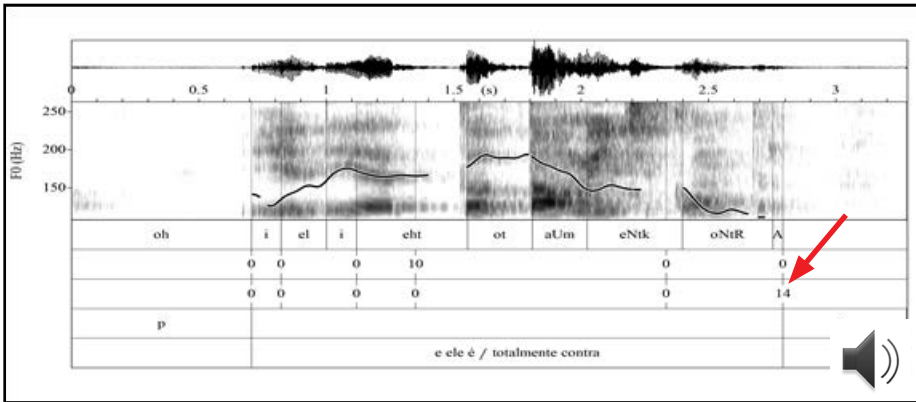
5th – Orthographic transcription tier.

(1) C-ORAL-BRASIL I, bfamnm24

e ele é/ totalmente contra //
 and he is totally against
 ‘And he is totally against it’

In this example, the terminal boundary occurs after the word “contra”. This utterance is formed by two tone units separated by the non-terminal boundary after the word “é”. The utterance on Example 1 ends with a silent pause, a falling pitch contour and lengthening of the pre-boundary V-V unit (Figure 1).

FIGURE 1 – Example 1 soundwave, spectrogram, pitch contour and textgrid



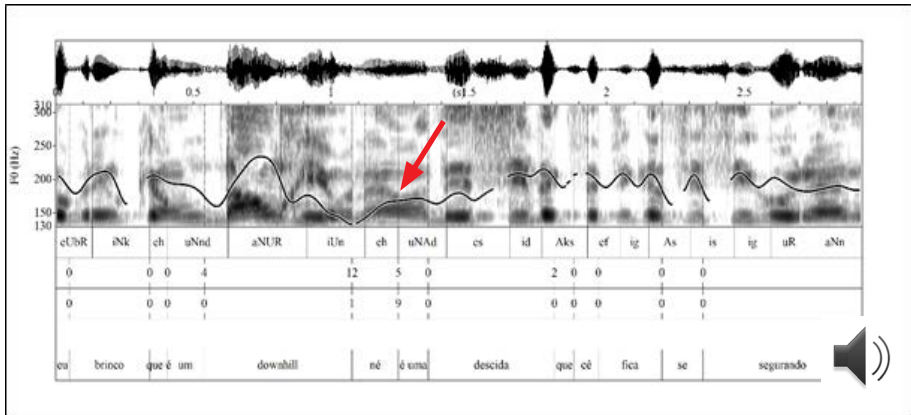
Other occurrences are not so prototypical in terms of parameter signaling, but still have a high prominence, as shown in Example 2. The boundary after the discourse marker “né” was perceived by all 14 annotators, but they were not in total agreement regarding if it was a terminal or non-terminal boundary. In Figure 2, it is possible to see that the boundary indicated by the red arrow does not present much of the prototypical features associated with boundaries (such as pitch reset, falling tone, pause), but 9 out of 14 listeners have identified it as a terminal boundary.

(2) C-ORAL-BRASIL II, bmidmn01

eu brinco que é um downhill / né // é uma descida que
 I joke that is a downhill DISC is a fall that
 cê fica se segurando
 you keep REFL holding

‘I joke that it’s like a downhill / you know // It’s a fall where you keep holding yourself’

FIGURE 2 – Example 2 soundwave, spectrogram, pitch contour and textgrid



From data inspection, we observed that terminal boundaries are usually highly prominent. Even so, it is not possible to distinguish a unifying prosodic description for boundaries that signal terminality. It could be argued that this is possibly related to the fact that an utterance may express different illocutive contents, prosodically encoded in many ways. However, we highlight the fact that, regardless the type of unit delimited by the boundary, listeners can perceive a common quality among different types of utterance closures. So, even though there are many possible ways to express utterance terminality, it is reasonable to expect that there are some acoustic cues that lead to the perception of “conclusiveness”.

Another aspect to be considered refers to utterances that are “abandoned”. For example, when the speaker drops the ongoing utterance and decides to start over, with a new one. Or, in another example, when the speaker is interrupted in mid utterance by external forces (for example, a loud noise or other participants in the conversation). In both cases, we

have the disruption of the utterance, which should be considered “closed”, but which is obviously not “concluded”.

This type of disfluency is also highly prominent for listeners, who usually have no doubt about the presence of a boundary. It is a very common phenomenon in spontaneous speech and implies extra challenges for an automatic recognition of prosodic boundaries, since these situations are not intentional. That means that there is no cognitive planning involved in the linguistic encoding of such events, hence, there is probably not a unifying set of prosodic parameters that indicate utterance interruption.

2.2 Non-terminal boundaries

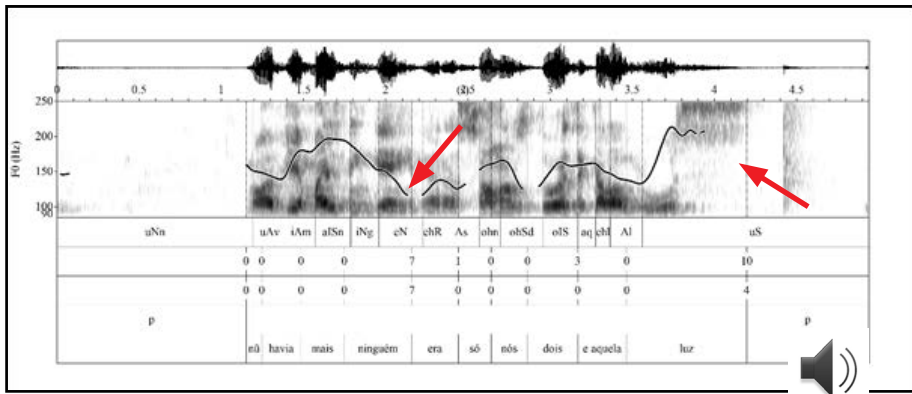
The non-terminal macrotype refers to prosodic boundaries that parse the utterance into smaller tone units. These boundaries are usually referred to in the literature as “continuative” boundaries. We prefer not to adopt this terminology, since prosodic boundaries that present a clear signal of discourse continuity are just one of the subtypes of non-terminal boundaries. Many prosodic boundaries do not carry a positive sign of continuity, but at the same time seem to lack a positive sign of utterance conclusion.

Example (3) presents two non-terminal boundaries: the first with a falling pitch after the word *ninguém* (“nobody”), usually associated with utterance finality; and the second with a rising pitch, after the word *luz* (“light”), usually associated with utterance continuation (Figure 3).

- (3) C-ORAL-BRASIL I, bfamnm11
nũ havia mais ninguém / era só nós dois e aquela luz /
 NEG there.was else anybody was just we two and that light
 ‘there wasn’t anybody else / it was just the two of us and that light /’

For the first boundary, the annotators were divided in relation to the nature of the boundary: 7 annotators identified it as a non-terminal and 7 as a terminal boundary, where as for the second boundary, 10 out of 14 annotators in our study identified it as a non-terminal boundary. That shows that the annotators have weighted different parameters in deciding as for boundary type and that pitch contour alone is not a sufficient predictor for boundary type distinction. Figure 3 shows both boundaries, indicated by red arrows.

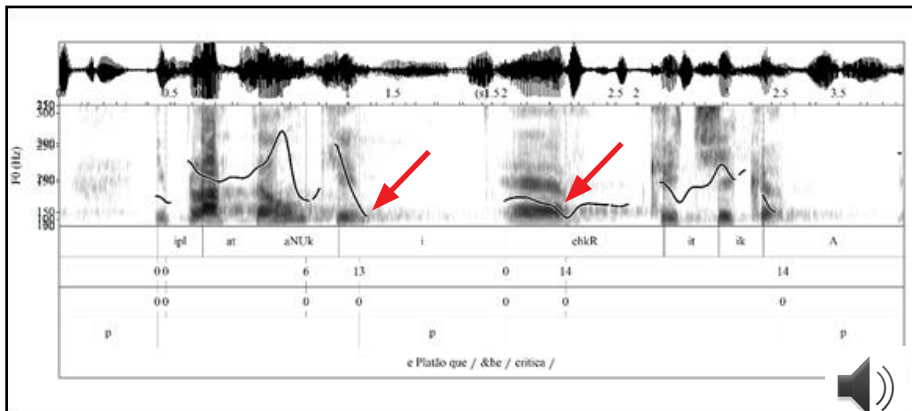
FIGURE 3 – Example 3 soundwave, spectrogram, pitch contour and textgrid



Example 4 and Figure 4 illustrate another type of non-terminal boundary, associated with utterance continuity. In this case, we have a filled pause delimited by two prosodic boundaries, indicated by the arrows (Figure 4).

- (4) C-ORAL-BRASIL II, bnatmn01
 e Platão que/ &he / critica /
 and Plato that / FILLER/ criticizes
 ‘and Plato that / eh / criticizes /’

FIGURE 4 – Example 4 soundwave, spectrogram, pitch contour and textgrid

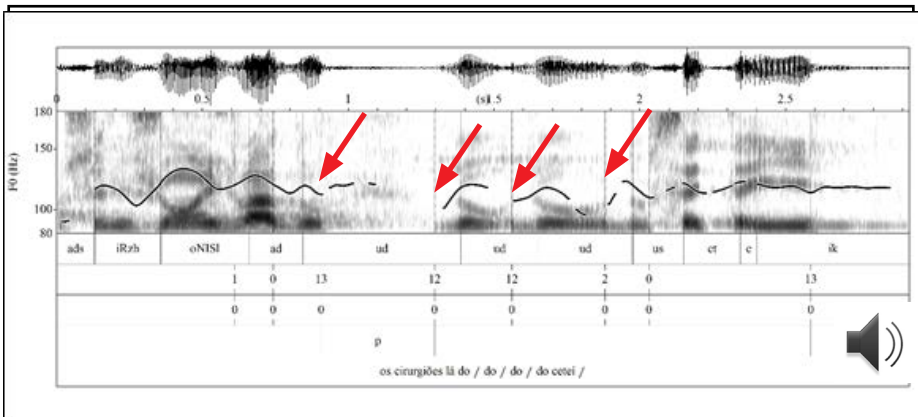


Pauses always indicate a disruption of the utterance and are a highly relevant indicative of boundary. However, they can be associated with either one of the two boundary macrotypes, terminal or non-terminal.

In example 5 we present another type of discourse disfluency, characterized by the lexical correction and/or lexical repetition of one or more items. This phenomenon is referred in this paper as “retracting” and is related to self-regulation in speech production, and it is usually formed by a single phonetic syllable. The acoustic features related to this type of non-terminal boundary make it challenging to model, since, similarly to utterance interruptions, they indicate a disfluency in speech and most likely are not realized through a consistent set of prosodic parameters.

- (5) C-ORAL-BRASIL II, bnatmn02
os cirurgiões lá do / do / do / do cetei /
 the surgeons DISC from.the from.the from.the from.the ICU
 ‘the ICU surgeons /’

FIGURE 5 – Example 5 soundwave, spectrogram, pitch contour and textgrid



We observe different instances of prosodic boundaries with acoustic correlates that differ from the categories they are usually associated with. Therefore, prosodic boundaries do not appear to be discrete categories, but rather partially stable instances, which are signaled through the variation of many acoustic parameters, as proposed by Barth-Weingarten (2016). Thus, the first step in the study of prosodic boundaries must consist in describing

the two more perceptually stable macrotypes, terminal and non-terminal, and then refine the study from that starting point.

Examples like the ones discussed here justify the hypothesis of the existence of more than two types of boundaries. They also explain the fact that some prosodic boundaries are highly prominent and perceived by (almost) all people, while others are not. Our operational hypothesis is that boundaries that are perceived by a higher number of people have more prototypical acoustic correlates, which are used more frequently in the language to signal terminality or non-terminality.

3. Methods

We extracted 7 excerpts of monological speech from Brazilian Portuguese spontaneous speech corpora C-ORAL-BRASIL I (RASO; MELLO, 2012) for informal speech and C-ORAL-BRASIL II (RASO; MELLO, in preparation) for formal and TV speech. The sample comprises a total of 1,339 words and 8 minutes and 39 seconds of male voices (Table 1).

TABLE 1 – Sample description

	Context	Gender	File ID	Time	Words
1	Informal	Male	bfammn11	01'11''	189
2	Informal	Male	bfammn24	00'58''	151
3	TV	Male	bmidmn01	01'23''	212
4	TV	Male	bmidmn02	01'21''	238
5	TV	Male	bmidmn03	01'07''	183
6	Formal	Male	bnatmn01	01'30''	205
7	Formal	Male	bnatmn02	01'09''	161
Total				08'39''	1339

We chose to perform this study using only the male monological speech because fundamental frequency differs a lot between men and women, and so we wanted to exclude the gender variable in this study. The methodological procedures are described in the following sections.

3.1. Data preparation

Each excerpt was independently segmented by fourteen annotators. Subjects were given the audio and transcript files with no punctuation or annotation besides turn separation and speaker identification. Annotators were asked to add mark-ups to the transcripts corresponding to their perception of prosodic boundaries, using the following symbols: single slash (/) for non-terminal boundaries and double slashes (//) for terminal boundaries. All subjects had already had some experience in prosodic segmentation of speech. Transcripts of all annotators were aligned word by word and the total number of annotators that signaled each position to the right of a word as a boundary was taken into account.

It is important to stress that different annotators may assign different boundary types to the same datum (see Example 3). For that reason we counted each boundary type separately. For this study, we decided that the model should consider as a boundary position every occurrence where at least 7 annotators (50%) signaled it as a terminal or a non-terminal boundary. That is, for the terminal boundary model, 7 or more annotators must have signaled the position as a terminal boundary; and the same for the non-terminal boundary model.

Additionally, after some initial tests, we decided to eliminate from the sample, all instances of non-terminal boundaries following retracting, filled pauses and the word “né”, given the high number of classification errors in those contexts.

Table 2 shows the total number of perceived boundaries in the sample.

TABLE 2 – Frequency of terminal and non-terminal boundaries perceived by at least 7 annotators

Boundary macrotype	Frequency	%
Terminal	70	24
Non-terminal	225	76
Total	295	100

In the next phase, all speech excerpts were annotated in Praat (BOERSMA; WEENINK, 2015) by creating a Textgrid with 5 tiers: an interval tier for Vowel to Vowel (V-V) broad phonetic transcription

(ASCII characters); a point tier for the number of subjects that identified each point as a non-terminal boundary (range 0-14); a point tier for the number of subjects that identified each point as a terminal boundary (range 0-14); an interval tier for silent pauses; an interval tier for orthographic transcription.

V-V units comprise the time between the onset of a vowel up to the onset of the next vowel and represent a phonetic syllable. V-V segmentation is adopted instead of a (phonological) syllabic segmentation because phonetic syllables represent more accurately the rhythmic structure of utterances (BARBOSA, 1996, 2006).

3.2. Acoustic parameters and data extraction

Based on literature review, a set of acoustic parameters was defined, to determine which parameters are better boundary correlates. Acoustic parameters are divided into five classes: a) speech rate and rhythm; b) standardized V-V duration; c) fundamental frequency (F0); d) intensity; e) silent pause.

Acoustic analysis considers each boundary in its surrounding context, and prosodic boundaries will always coincide with boundaries of phonological words. Thus, the context for analysis is defined as 21 V-V units centered at a given phonological word boundary. This includes positions signaled by annotators as boundaries or non-boundaries. That means two windows of analysis, one including 10 V-V units to the left and one with 10 V-V units to the right of a position in analysis plus the V-V unit that starts at the current position.

Table 3 shows a summary of the measurements extracted for prosodic analysis, divided into global and local. Global measurements are calculated considering the values from left and right windows, plus the difference between those values at a phonological word boundary position. Local values are calculated for every single V-V unit inside the left and right windows.

TABLE 3– Summary of acoustic parameters

Class	Type	Measurement
Speech rate and rhythm	Global	Rate of V-V units per second (right window context, left window context and difference) Rate of non-salient V-V units per second
	Local	Mean of smoothed z-score (adjacent right context, adjacent left context and difference)
Standardized segment duration	Global	Mean of smoothed z-score (right window context, left window context and difference)
		Standard deviation of smoothed z-score (right window context, left window context and difference)
		Skewness of smoothed z-score (right window context, left window context and difference)
		Peak rate of smoothed z-score (right window context, left window context and difference)
Fundamental frequency	Local	F0 median for each V-V (left and right V-Vs in window and difference at window center) in semitones re 1 Hz
		First derivative of F0 median for each V-V unit (left and right V-Vs in window and difference at window center) in semitones re 1 Hz/s
	Global	Mean of F0 medians (right window context, left window context and difference) in semitones re 1 Hz
		Standard deviation of F0 medians (right window context, left window context and difference) in semitones re 1 Hz
		Skewness of F0 medians (right window context, left window context and difference)
		Mean of F0 median first derivative (right window context, left window context and difference) in semitones re 1 Hz/s
		Standard deviation of F0 median first derivative (right window context, left window context and difference) in semitones re 1 Hz/s
Peak rate of smoothed F0 peaks per second (right window context, left window context and difference)		
Intensity	Local	Mean spectral emphasis for V-V unit at window center in dB
	Global	Mean spectral emphasis (right window context, left window context and difference) in dB
Pause	Local	Pause presence (0 = absence or 1 = presence)
		Pause duration in seconds

Data extraction was performed through *BreakDescriptor* (BARBOSA, 2016), a Praat script developed from *ProsodyDescriptor* (BARBOSA, 2013). *BreakDescriptor* calculates and extracts acoustic data from every V-V unit (phonetic syllable) of the analysis context, which comprises 10 units to the left and 10 units to the right of the phonetic syllable under analysis plus the phonetic syllable itself. That comprises a total of 111 acoustic measurements for each position, according to the variables described in Table 3.

3.3 Evaluation of classification methods

Our goal is to arrive to a set of acoustic parameters that can identify the chance that any given phonological word boundary corresponds to a terminal prosodic boundary, a non-terminal prosodic boundary or none. Thus, we search for a model that assigns a weight to each acoustic parameter and ensures the greatest possible discrimination between any of the two macrotypes of prosodic boundaries and the absence of prosodic boundaries.

For this purpose, we tested two classification methods: Random Forest (RF) and Linear Discriminant Analysis (LDA). These methods of statistical classification were used to obtain hierarchical classification models based on the observation of the predictor variables, in this case acoustic parameters (Table 3). This process makes it possible to identify the combination of measurements and weights that can best explain the perceptual segmentation performed by human annotators. LDA and Random Forest are two statistical techniques that result in different models, While LDA calculates association through linear regression, Random Forest uses decision trees, also called decision nodes.

Calculations were performed with the R environment for statistical computing (R CORE TEAM, 2017). The LDA method is part of the *MASS* package (VENABLES; RIPLEY, 2002) – function *lda()*. The RF method is found in the *randomForest* package (LIAW; WIENER, 2002) – function *randomForest(x, ntree=100, proximity=TRUE)*.

For the evaluation of both methods we verified results for both a training stage and a test stage. During the training stage, the classification method infers weights of predictor variables and performs a multivariate analysis of data, to arrive at statistical correlations between predicted (boundary presence or absence) and predictor variables (acoustic parameters) for all groups. The test stage evaluates the effectiveness of

the classification method in distinguishing the groups of boundary vs. non-boundary. We created two separate samples, one for training and one for testing. The training set consisted of a random selection of 70% of the V-V units in our data, whereas the test set consisted of the remaining 30% of the V-V units.

For both classification methods, we considered the presence and the absence of a certain boundary type, for both, terminal and non-terminal boundaries, building a separate model for each. Thus, in the terminal boundary model, absence of boundary includes also the instances of non-terminal boundaries; and in the non-terminal boundary model the absence of boundary includes also the instances of terminal boundaries.

We also consider the predictive power of LDA and RF. The prediction shows hits and false alarms for the dataset. After an initial evaluation phase, the LDA method presented the best results for both boundary types, producing a better match to the perceptual segmentation. Therefore, the LDA method was further refined, in order to improve the performance of the classifier as well as to reduce to a minimum the number of variables used for classification.

3.4 LDA refinement

LDA refinement consisted in identifying the most and least relevant variables among the 111 acoustic parameters collected by *BreakDescriptor*. The gradual elimination of parameters allowed us to achieve the highest percentage of hits for boundary presence and, the lowest percentage of false alarms in points perceived as absent of boundary as well as a minimum set of predictors, which allows to reduce the window extension around each predicting position.

For the refinement, we also split the set into a training set with 70% of random positions and a test set with the remaining 30%.

The LDA model refinement was carried out in two phases. In the first phase, the measurements extracted by *BreakDescriptor* were gradually removed from each model by discarding the ones with the lowest weights. In the second phase, measurements were excluded from the model based on the phonetic phenomena they represent, based on literature review. Thus, the less relevant phonetic phenomena were eliminated. This process aimed at reducing the “noise” in the models, increasing the proportion of hits and reducing the proportion of false alarms with a reduced set of acoustic predictors.

Finally, we investigated the hypothesis that the non-terminal boundaries in the dataset represent different boundary sub-types, signaled by different groupings of acoustic parameters. For this, we did not perform training and testing. Instead, in order to maximize our available sample, we used the entire dataset, except all instances of boundaries identified by 7 or more annotators as terminal boundaries. We then performed a cluster analysis to identify possible groups of similar non-terminal boundaries. Clusters were calculated using the complete linkage method, through R environment for statistical computing (R CORE TEAM, 2017), with the function *hclust()*. The dissimilarity matrix for the cluster analysis was calculated using the Euclidean method with the function *dist()* from a table of correlations of parameters obtained by Pearson’s coefficient, with the function *cor(x, method=“pearson”)^2*. All these functions belong to the *stats* package included in R core.

4. Results

4.1 RF and LDA Evaluation

Evaluation of models generated by RF and LDA classification methods took into consideration all 111 acoustic parameters as predictor variables for presence or absence of terminal and non-terminal boundaries. Table 4 shows absolute values for identification of boundaries. These results show that the LDA model identified a higher number of terminal boundaries, and was also able to identify the absence of terminal and non-terminal boundaries in a higher number of occurrences.

TABLE 4 – Evaluation of RF and LDA, absolute frequency of boundary identification

Boundary	RF		LDA	
	Terminal	Non-terminal	Terminal	Non-terminal
Presence	47	185	75	142
Absence	785	646	1076	1010

Based on these results and the data from the perceptual annotation of prosodic boundaries (Table 2), we calculated the predictive power of the models generated by each classification method. The predictive power establishes the percentage of hits and false alarms for each boundary

macrotype. A hit indicates that the statistical model was able to identify a boundary that was perceived as such by at least 50% of human annotators. A false alarm means that the model predicts a boundary where human annotators did not perceive one.

We obtained the following results:

- a) **Terminal boundaries:** RF predicted 28% of terminal boundaries correctly, whereas it has only 1% of false alarms. LDA, on the other hand, has 57% of terminal boundaries hits and 2% of false alarms.
- b) **Non-terminal boundaries:** RF predicted 19% of terminal boundaries correctly, whereas it has only 6% of false alarms. LDA, on the other hand, has 38% of terminal boundaries hits and 5% of false alarms.

Based on this, we proceed with the refinement of the models generated by the LDA classifier.

4.2 Refining the LDA model for terminal boundaries

The first model included all 111 acoustic parameters extracted by *BreakDescriptor*. Frequency of terminal boundaries and the model predictive power are presented in Table 5. For the model with all 111 parameters, the LDA classifier produces 76% of hits and 24% of false alarms for terminal boundaries. LDA model showed 97.4% correct prediction for the absence of terminal boundaries.

TABLE 5 – Frequency of boundary identification and predictive power of model for terminal boundaries with 111 acoustic parameters

Terminal Boundary	Frequency	% Correct	% Wrong
Presence	38	76	24
Absence	759	97.4	2.6

We progressively removed the least relevant acoustic parameters based on phonetic criteria. The model that presented the best results for terminal boundary classification used 20 of the 111 parameters. Table 6 shows the results of performance of this final model for terminal boundaries. The model reached a convergence with human annotation of 80% for boundary presence and 92% for boundary absence.

TABLE 6 – Frequency of boundary identification and predictive power of model for terminal boundaries with 20 acoustic parameters

Terminal boundary	Training			Test		
	Freq.	% Correct	% Wrong	Freq.	% Correct	% Wrong
Presence	45	80	20	25	80	20
Absence	837	95.2	4.8	319	92	8

The set of parameters that constitute the model for terminal boundaries is listed in Table 7. Results show that pauses are the most relevant parameters for classifying a boundary as terminal. The next parameters indicate changes in pitch direction and pitch reset, followed by pre-boundary syllabic lengthening and changes in speech rate. Finally, the relative intensity in the pre-boundary syllable also contributes to the identification of terminal boundaries.

TABLE 7 – Parameters for identification of terminal boundaries according to statistical weight

Parameter class	Abbrev.	Weight	Global/local parameter measurement
Pause	psdur	2.641	Pause duration after V-V unit.
	psp	1.948	Pause presence after V-V unit.
Fundamental frequency	f0meddloc	0.329	First derivative of F0 median: difference between V-V at boundary and first V-V to the right.
	df0medr1	0.264	First derivative of F0 median for 1st V-V unit on right window.
	df0medl	0.257	Mean of F0 median first derivative on the left windows.
	sddf0d	0.157	Standard deviation of first derivative of F0 median: difference between right and left V-V unit.
Normalized duration of syllabic segments	prd	0.101	Peak rate of smoothed z-score: difference between right and left windows.
Fundamental frequency	sdf0l	0.091	Standard deviation of F0 medians on left window.
	df0medl10	0.066	First derivative of F0 median for 10th V-V unit on left window.
	f0rl	0.033	Peak rate of smoothed F0 peaks per second on the left windows.
	df0meddloc	0.032	First derivative of F0 median: difference between 1st V-V unit on right window and V-V unit at boundary point.
	f0medd	0.029	Mean of F0 medians: difference between right and left windows.

Normalized duration of syllabic segments	zl10	0.028	Mean of smoothed z-score for 10th V-V unit on the left window.
Fundamental frequency	skf0d	0.025	Skewness of F0 medians: difference between right and left windows.
Normalized duration of syllabic segments	mzd	0.015	Mean of smoothed z-score: difference between right and left windows.
Fundamental frequency	skdf0d	0.011	Skewness of F0 first derivative medians: difference between right and left windows.
Normalized duration of syllabic segments	SDzl	0.010	Standard deviation of smoothed z-score: difference between V-V units on left window.
Speech rate and rhythm	ard	0.003	Rate of non-salient V-V units per second: difference between right and left windows.
Normalized duration of syllabic segments	zdlc	0.001	Mean of smoothed z-score: difference between first V-V unit on right window and V-V unit at boundary point
Intensity	emphl	0.001	Mean spectral emphasis on left window

The model for terminal boundaries is consistent with the description of prototypical “conclusive” boundaries found in the literature. The model presents a clear hierarchy of acoustic parameters and also describes their relative importance. Additionally, it highlights the relevance of global measurements. That reinforces the notion that prosodic boundaries are not a localized phenomenon, but are related to the prosodic structuring of the utterance.

4.3 Refining the LDA model for non-terminal boundaries

The first model for non-terminal boundaries included all 111 acoustic parameters extracted by *BreakDescriptor*. Frequency of non-terminal boundaries and the model predictive power are presented in Table 8. The LDA classifier produces 39% of hits and 61% of false alarms for non-terminal boundaries. LDA model showed 94.9% correct prediction for the absence of terminal boundaries.

TABLE 8 – Frequency of boundary identification and predictive power of model for non-terminal boundaries with 111 acoustic parameters

Non-terminal Boundary	Frequency	% Correct	% Wrong
Presence	179	39	61
Absence	618	94.9	5.1

In comparison with the first model for terminal boundaries, this result indicates a lower predictive power, with a higher number of false boundary identification. Non-terminal boundaries seem to be signaled by more diverse parameters that appear not to fit into a single group, thus, they present greater variety of sub-types than terminal boundaries, thus corroborating the notion of boundary macrotypes.

By progressive elimination of boundaries according to phonetic criteria, we arrived at a second model with 9 parameters. Frequency of boundary identification and predictive power are presented in Table 9. We observe little improvement when comparing tests results in Tables 8 and 9. Although boundary hit frequency is now 50%, the number of false alarms decreased 11% in comparison with the previous model.

TABLE 9 – Frequency of boundary identification and predictive power of model for non-terminal boundaries with 9 acoustic parameters

Terminal boundary	Training			Test		
	Freq.	% Correct	% Wrong	Freq.	% Correct	% Wrong
Presence	60	37.2	62.8	32	50	50
Absence	685	95.1	4.9	257	92.8	7.2

Since the model could not be further improved, we decided to investigate the hypothesis that this dataset represents more than one sub-type of non-terminal boundary. We used the entire dataset for these last rounds of refinement (instead of using 70% for training and 30% for testing as in previous phases) and excluded instances of terminal boundaries (see Methods).

In the first round, we tested our dataset with Model 1 (9 parameters). On the next round, we took all instances of non-terminal boundaries and boundary absence that were not identified correctly by Model 1 to generate Model 2 (10 parameters). We applied the same procedure one more time, taking all instances of non-terminal boundaries and boundary absence that were not identified by Model 2 to generate Model 3 (8 parameters). These 3 models accounted for 220 (out of 225, see Table 2) of non-terminal boundaries in our dataset.

Table 10 shows the frequency of identification of prosodic boundaries and also the predictive power for the three models. With

this method, we increased the hits and decreased the number of false alarms in all three new models in comparison to the preceding models. The three new models capture, with more detail, the differences among distinct subtypes of non-terminal boundaries.

Model 1 identified more boundaries (69% of the total boundaries automatically assigned by all three models), but, at the same time, it had the worst convergence with human annotators, with 68% hits and 32% of false alarms, and also the worst performance identifying boundary absence. Model 2 identified 57 boundaries (26%) and had the best convergence with human annotators, with 78% of hits and 22% of false alarms. Model 3 identified very few boundaries (5%) and has the best convergence for boundary absence identification.

TABLE 10 – Frequency of boundary identification and predictive power of 3 models for non-terminal boundaries

Model	Boundary presence				Boundary absence			
	Freq.	%	% Correct	% Wrong	Freq.	%	% Correct	% Wrong
Model 1 – 9 parameters	152	69	68	32	125	58	78	22
Model 2 – 10 parameters	57	26	78	22	52	24	80	20
Model 3 – 8 parameters	11	5	69	31	37	17	88	12

Table 11 presents the list of prosodic parameters selected by each model. The first column shows the rank for all parameters. For each model, the first column indicates the abbreviations assigned for predictors and the statistical weight for the measurement; the second column has a full description of the measurement calculated for each parameter. All models are composed by a different set of acoustic measurements.

TABLE 11—Models for identification of non-terminal boundaries – parameters ranked by statistical weight

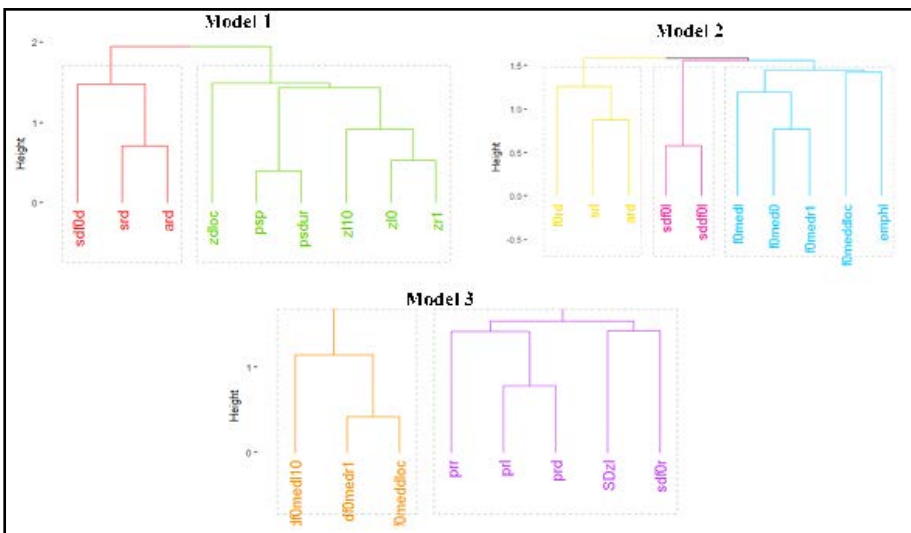
Rank	Model 1 – 9 parameters	Model 2 – 10 parameters	Model 3 – 8 parameters
1 st	zl0 4.5 Mean of smoothed z-score of V-V unit at boundary point	srl 0.72 Rate of V-V units per second on the left window	prl 151.6 Peak rate of smoothed z-score on left window
2 nd	zr1 4.4 Mean of smoothed z-score 1st right	sddf0l 0.63 Standard deviation of F0 median first derivative on left window	prd 150.6 Peak rate of smoothed z-score - difference between right and left window
3 rd	zdloc 4.2 Mean of smoothed z-score - difference between 1st V-V on right window and and V-V unit at boundary point	sdf0l 0.47 Standard deviation of F0 medians on left window	pr 149.5 Peak rate of smoothed z-score on right window
4 th	psp 2.6 Pause presence	ard (*) 0.45 Rate of non-salient V-V units per second - difference between right and left windows	sdf0r 0.5 Standard deviation of F0 medians on right window
5 th	psdur 2.3 Pause duration	f0medl 0.37 Mean of F0 medians left window context	SDzl 0.3 Standard deviation of smoothed z-score on left window
6 th	ard (*) 0.3 Rate of non-salient V-V units per second - difference between right and left windows	f0rd 0.21 Peak rate of smoothed F0 peaks per second difference of right and left windows	df0medr1 0.3 First derivative of F0 median for 1st V-V unit on the right
7 th	srd 0.3 Rate of V-V units per second - difference between right and left windows	f0meddloc 0.10 F0 median - difference between last V-V unit on the left window and first unit on the right	df0medl10 0.2 First derivative of F0 median for 10st V-V unit on the left
8 th	sdf0d 0.2 Standard deviation of F0 medians - difference between right and left windows	f0med0 0.09 F0 median of V-V unit at boundary point	df0meddloc 0.1 F0 median - difference between first V-V unit on the right window and V-V unit at boundary point
9 th	zl10 0.2 Mean of smoothed z-score for 10th V-V unit on the left window	f0medr1 0.05 F0 median of V-V unit at 1st V-V unit on the right	
10 th		emphl 0.01 Mean spectral emphasis on the left window	

(*) measurement present in more than one Model.

Model 1 identifies boundaries signaled through parameters related mainly to the organization of phonetic syllables in time, like duration of phonetic syllables (z10, zr1, zdloc, z110), presence and duration of pause (psp, psdur) and speech and articulation rates (srd, ard), with only acoustic parameter related to pitch movements (sdf0). The most relevant measurements in Model 1 regard local, duration related, parameters. Model 2 identifies non-terminal boundaries based on pitch excursion, reset and prominence (sddf0l, sdf0l, f0medl, f0rd, f0medloc, f0med0, f0medr1), speech and articulation rates (srl, ard) and, in a lower degree, intensity (emphl). Global pitch parameters are the most relevant for Model 2. Model 3 identifies boundaries signaled mainly through saliences in syllabic durations (prl, prd, prr, SDzl) and local variations in pitch (sdf0r, dfmedr1, dfmedl10, dfmeddloc). Global duration parameters are the more relevant in Model 3.

Figure 5 shows the clusters obtained for each model, acoustic parameters are identified by the abbreviations shown in Table 11. Clusters allow us to detect subtypes of boundaries and a more detailed view of the relevant parameters for boundary identification through subgroups of boundary predictors. For each subgroup in each model, the division of parameters mostly falls into the broad classes of prosodic parameters: speech rate and rhythm; segment duration; fundamental frequency, intensity and silent pause.

FIGURE 6 – Clusters of parameters for non-terminal prosodic boundaries



Model 1 presents 2 main groups and 6 subgroups of parameters. On the left side, the first main group aggregates global parameters with lower weight that indicate speech and articulation rates and pitch variation. On the right side, the second main group presents local parameters with higher weight related to syllabic lengthening.

Model 2 has three main groups and 6 subgroups. The left main group contains global parameters that indicate speech and articulation rates and pitch prominences (related to pitch accents). The center main group includes two global parameters that indicate pitch variations. The right main group combines one local and three global parameters related to pitch movements and a global parameter of intensity.

Model 3 consists of 2 main groups and 5 subgroups. The first main group on the left aggregates local parameters that indicate abrupt changes in pitch. Finally, the last group on the right is composed only by global parameters. It combines four duration parameters related to duration saliences and rhythmic variations and one parameter indicating pitch variations.

The clusters corroborate the notion that prosodic boundaries are a complex and granular phenomenon, that is, the non-terminal category encompasses boundaries signaled by different sets of acoustic parameters, which probably correlate with different boundary sub-types.

5. Final remarks

The results indicate that the Linear Discriminant Analysis classifier provides better models for the terminal and non-terminal boundary macrotypes. After the refinement of the model generated by LDA, we were able to attest the adequacy of this method. Despite the number of false alarms, the models represent a good fit regarding the decisions made by annotators in our dataset.

Our results point to a higher degree of predictive performance related to terminal boundaries. The resulting model has a higher number of hits and fewer mistakes in relation to non-terminal boundaries. At least in the database used in this study, signalization of utterance conclusion seems to be more typified, while signalization of boundaries from the non-terminal macrotype appears to be more stratified. Another question that arises from this stratification is if there are linguistic or perceptual correlates for the different boundary Models and its subgroups. Further tests with more diverse data are needed to verify these hypotheses.

Other line of investigation refers to the analysis of errors for each model. These instances could reveal finer details regarding segmentation. Do these instances represent annotators ambiguity in boundary identification? Or are there other non-terminal boundary sub-types that are just under-represented in the sample? How many of these errors are due to disfluencies (interruption, time-taking, retracting) and is it possible to model those phenomena? Understanding the contexts where the model does not fit the human annotation would be useful to produce better models.

This research achieved its proposed goal to present models for the prediction of prosodic boundaries, based on spontaneous speech data. Next stages of this research would involve an increase in the database, so more extensive testing can be performed to produce robust models that can be used for the automatic segmentation of speech.

Authors' Contributions

Both authors contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript.

References

AUER, P. Zum Segmentierungsproblem in der Gesprochenen Sprache. *InLiSt - Interaction and Linguistic Structures*, Freiburg, v. 49, p. 1-19, Nov. 2010. Available from: <<http://www.inlist.uni-bayreuth.de/issues/49/InList49.pdf>>. Access on: 5 Dec. 2017.

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

BARBOSA, P. A. At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration. In: ETRW ON SPEECH PRODUCTION MODELING: FROM CONTROL STRATEGIES TO ACOUSTIC, 1., 1996, Autrans. p. 85-88. Available from: <http://www.isca-speech.org/archive_open/spm_96/sps6_085.html>. Access on: 5 Dec. 2017.

BARBOSA, P. A. *BreakDescriptor*. Script para o PRAAT. [Computer program]. 2016.

BARBOSA, P. A. *Incursões em torno do ritmo da fala*. Campinas: Pontes; Fapesp, 2006.

BARBOSA, P. A. Prominence-and boundary-related acoustic correlations in Brazilian Portuguese read and spontaneous speech. In: BARBOSA, P. A.; MADUREIRA, S.; REIS, C. (Ed.). *Speech Prosody*. Campinas: ISCA, 2008. p. 257-260. Available from: <<http://aune.lpl.univ-aix.fr/~sprog/sp2008/papers/id060.pdf>>. Access on: 5 Dec. 2017.

BARBOSA, P. A. Conhecendo melhor a prosódia: aspectos teóricos e metodológicos daquilo que molda nossa enunciação. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 20, n. 1, p. 11-27, 2012.

BARBOSA, P. A. Semi-automatic and automatic tools for generating prosodic descriptors for prosody research. In: BIGI, B.; HIRST, D. (Eds.). *Proceedings of the Tools and Resources for the Analysis of Speech Prosody*. Aix-en-Provence: Laboratoire Parole et Langage, 2013. v. 13, p. 86-89. Available from: <<http://www.lpl-aix.fr/~trasp/Proceedings/19874-trasp2013.pdf>>. Access on: 22 Dec. 2015.

BARTH-WEINGARTEN, D. *Intonation Units Revisited: Cesuras in talk-in-interaction*. Amsterdam: John Benjamins, 2016.

BATLINER, A. *et al.* The Prosodic Marking of Phrase Boundaries: Expectations and Results. In: RUBIO AYUSO, A. J.; LOPEZ SOLER, J. M. (Org.). *Speech Recognition and Coding: New advances and Trends*. Berlin: Springer, 1995. v. 147, p. 89-92.

BIRKNER, K. Relative Konstruktionen zur Personenattribution. In: GÜNTHER, S.; WOLFGANG, I. *Konstruktionen in der Interaktion*. Berlin: Mouton de Gruyter, 2006. p. 205-238.

BLAAUW, Eleonora. The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication*, Elsevier, v. 14, n. 4, p. 359-375, 1994. Available at: <<http://www.sciencedirect.com/science/article/pii/0167639394900280>>. Access on: 10 Apr. 2015.

BOERSMA, P.; WEENINK, D. *Praat: doing phonetics by computer*. 2015. Available from: <<http://www.praat.org/>>. Access: 2 dec. 2015

BOLINGER, D. Around the edges of language. In: BOLINGER, D. (Ed.). *Intonation: Selected Readings*. Harmondsworth: Penguin, 1972. p. 19-29.

BOSSAGLIA, G. Effects of speech rhythm on spoken syntax A corpus-based study on Brazilian Portuguese and Italian. *CHIMERA: Romance Corpora and Linguistic Studies*, Madri, v. 2, n. 3, p. 265-285, 2016.

- BROWN, G. *et al. Questions of Intonation*. London: Croom Helm, 1980.
- BYBEE, J. *Language, usage and cognition*. Cambridge: Cambridge University Press, 2010.
- CHAFE, W. L. *Discourse, consciousness and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago, 1994.
- COLE, J.; SHATTUCK-HUFNAGEL, S.; MO, Y. Prosody production in spontaneous speech: Phonological encoding, phonetic variability, and the prosodic signature of individual speakers. *The Journal of the Acoustical Society of America*, New York, v. 128, n. 4, p. 2429, 2010.
- COOPER, W. E.; PACCIA-COOPER, J. *Syntax and speech*. Cambridge/MA: Harvard University Press, 1980.
- COUPER-KUHLEN, E. Prosodic Cues of Discourse Units. In: BROWN, Keith (Ed.). *Encyclopedia of Language & Linguistics*. Oxford: Elsevier, 2006. p. 178-182.
- CRESTI, E. *Corpus di Italiano parlato*. Firenze: Accademia della Crusca, 2000. v. 1.
- CRESTI, E. Syntactic properties of spontaneous speech in the L-AcT framework: data on Italian complement and relative clauses through the IPIC Data Base. In: RASO, T.; MELLO, H.; PETTORINO, M. (Ed.). *Spoken Corpora and Linguistic Studies*. Philadelphia; Amsterdam: John Benjamins, 2014.
- CRESTI, E.; MONEGLIA, M. Informational patterning theory and the corpus-based description of spoken language: The compositionality issue in the topic-comment pattern. In: MONEGLIA, M.; PANUNZI, A. (Ed.). *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*. Firenze: Firenze University Press, 2010. p. 13-45.
- CROFT, W. Intonation units and grammatical structure. *Linguistics*, De Gruyter, v. 33, n. 5, p. 839-882, 1995.
- CRUTTENDEN, A. *Intonation*. 2. ed. Cambridge: CUP, 1997.
- CRYSTAL, D. *Prosodic Systems and Intonation in English*. Cambridge: CUP, 1969.

DU BOIS, J. W.; CUMMING, S.; SCHUETZE-COBURN, S.; PAOLINO, D. (Ed.). *Santa Barbara Papers in Linguistics*. v. 4: Discourse Transcription. *Santa Barbara Papers in Linguistics*, Santa Barbara, v. 4, 224p., 1992.

DU BOIS, J. *Rhythm and Tunes: The notation Unit in the Structure of Dialogic Engagement*. Conference Prosody and Interaction. University of Potsdam, 2008.

FON, J.; JOHNSON, K.; CHEN, S. Durational patterning at syntactic and discourse boundaries in Mandarin spontaneous speech. *Language and Speech*, Kansas, v. 54, n. Pt 1, p. 5-32, 2011.

FOWLER, C. A. Segmentation of coarticulated speech in perception. *Attention, Perception & Psychophysics*, New York, v. 36, n. 4, p. 359-368, 1984.

FUCHS, S.; KRIVOKAPIC, J.; JANNEDY, S. Prosodic boundaries in German: Final lengthening in spontaneous speech. *The Journal of the Acoustical Society of America*, New York, v. 127, n. 3, p. 1851, 2010.

HALLIDAY, M. A. K. *Speech and Situation*. London: University College, 1965.

IZRE'EL, S. Intonation Units and the Structure of Spontaneous Spoken Language : A View from Hebrew. In: AURAN, C; BERTRAND, R; CHANET, C; COLAS, A; DI CRISTO, A; PORTES, C; REYNIER, A; VION, M. (Ed.) *Proceedings of the IDP05 International Symposium on Discourse-Prosody Interfaces*. Aix-en-Provence: 2011. Available from: <<http://aune.lpl.univ-aix.fr/~prodige/idp05/actes/izreel.pdf>>. Access at: 20 Nov. 2017.

KOHLER, K. J; PETERS, B.; WESENER, T. Interruption Glottalization in German Spontaneous Speech. *Proceedings of Disfluency in Spontaneous Speech*, DiSS01, 2001. p. 45-48. Available from: <http://www.isca-speech.org/archive_open/archive_papers/diss_01/dis1_045.pdf>. Access at: 20 Nov. 2017.

LIAW, A.; WIENER, M. Classification and Regression by randomForest. *The R News Journal*, [s.l.], v. 2, n. 3, p. 18-22, 2002. Available from: <<http://cran.r-project.org/doc/Rnews/>>. Access on: 10, Jan. 2018.

MITTMANN, M. M. *et al.* Utterance as the minimal pragmatic entity in spontaneous speech perception. In: CONFERÊNCIA LINGÜÍSTICA E COGNIÇÃO, V., 2010, Florianópolis. *Anais...* Florianópolis: Universidade Federal de Santa Catarina, 2010. Available from: <<http://www.nupffale.ufsc.br/lincognition/anais.htm>>. Access on: 20, Nov. 2017.

MITTMANN, M. M.; BARBOSA, A. An automatic speech segmentation tool based on multiple acoustic parameters. *CHIMERA. Romance Corpora and Linguistic Studies*, Madri, v. 32, p. 133-147, 2016.

MO, Y.; COLE, J.; LEE, E-K. Naïve listeners' prominence and boundary perception. In: BARBOSA, P. A.; MADUREIRA, S.; REIS, C. (Org.). *Speech Prosody*. Campinas: ISCA, 2008. p. 735-738. Available from: <http://www.isca-speech.org/archive/sp2008/papers/sp08_735.pdf>. Access on: 20, Nov. 2017.

MO, Y. Duration and intensity as perceptual cues for naïve listeners' prominence and boundary perception. In: BARBOSA, P. A.; MADUREIRA, S.; REIS, C. (Ed.). *Speech Prosody*. Campinas: ISCA, 2008. Available from: <http://www.isca-speech.org/archive/sp2008/sp08_739.html>. Access on: 20 Nov. 2017.

MONEGLIA, M.; CRESTI, E. C-ORAL-ROM: Prosodic boundaries for spontaneous speech analysis. In: KAWAGUCHI, Y.; ZAIMA, S.; TAKAGAKI, T. (Ed.). *Spoken Language Corpus and Linguistics Informatics*. Amsterdam; Philadelphia: John Benjamins, 2006. p. 89-112.

MONEGLIA, M. Units of Analysis of Spontaneous Speech and Speech Variation in a Cross-linguistic Perspective. In: KAWAGUCHI, Y.; ZAIMA, S.; TAKAGAKI, T. (Ed.). *Spoken Language Corpus and Linguistics Informatics*. Amsterdam; Philadelphia: John Benjamins, 2006. p. 153-179.

MONEGLIA, M. Spoken Corpora and Pragmatics. *Revista Brasileira de Linguística Aplicada*, Belo Horizonte, v. 11, n. 2, p. 479-519, 2011.

PETERS, B.; KOHLER, K. J.; WESENER, T. Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache. In: KOHLER, J.; KLEBER, F.; PETERS, B. (Ed.). *Prosodic Structures in German Spontaneous Speech* (AIPUK 35a). Kiel: IPDS, 2005. p. 143-184.

PIERREHUMBERT, J. B. *The Phonetics and Phonology of English Intonation*. 1980. 401 f. Thesis (PhD) – Dept. of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge/MA, 1980. Available from: <<http://hdl.handle.net/1721.1/16065>>. Access on: 20 Nov. 2017.

PIKE, K. L. *The intonation of American English*. Ann Arbor: University of Michigan, 1945.

R CORE TEAM (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. [Computer program] Available from: <<https://www.R-project.org/>>. Access on: 15 Dec. 2017.

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

RASO, T.; MITTMANN, M. M.; MENDES, A. C. O. O papel da pausa na segmentação prosódica de corpora de fala. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 23, n. 3, p. 883-922, 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*, Madri, v. 3, p. 221-249, 2016.

SANDERS, L. D.; NEVILLE, H. J. Lexical, Syntactic, and Stress-Pattern Cues for Speech Segmentation. *Journal of Speech, Language, and Hearing Research*, ASHA Association, v. 43, n. 6, p. 1301-1321, 2000.

SCHUETZE-COBURN, S.; SHAPLEY, M.; WEBER, E. G. Units of intonation in discourse: a comparison of acoustic and auditory analyses. *Language and Speech*, Kansas, v. 34, n. 3, p. 207-234, 1991.

SELKIRK, E. O. Comments on Intonational Phrasing in English. In: FROTA, S.; VIGARIO, M.; FREITAS, M. J. (Ed.). *Prosodies*. Berlin: Mouton de Gruyter, 2005. p. 11-58.

SWERTS, M.; COLLIER, R.; TERKEN, J. Prosodic predictors of discourse finality in spontaneous monologues. *Speech Communication*, Elsevier, v. 15, n. 1-2, p. 79-90, Out. 1994.

SWERTS, M. Prosodic features at discourse boundaries of different strength. *The Journal of the Acoustical Society of America*, New York, v. 101, n. 1, p. 514-521, 1997.

SZCZEPEK REED, B. Turn-final intonation in English. In: COUPER-KUHLEN, E.; FORD, C. (Ed.). *Sound Patterns in Interaction*. Amsterdam: John Benjamins, 2004. p. 97-118.

SZCZEPEK REED, B. Prosody, syntax and action formation: Intonation phrases as “action components”. In: BERGMANN, P. *et al.* (Ed.). *Prosody and Embodiment in Interactional Grammar*. Berlin: Mouton de Gruyter, 2012. p. 142-170.

TSENG, C.-Y. Y. *et al.* Fluent speech prosody: Framework and modeling. *Speech Communication*, Elsevier, Anais... jul. 2005. Available from: <<http://www.sciencedirect.com/science/article/pii/S0167639305000919>>. Access on: 26 May 2015.

TSENG, C.-Y.; CHANG, C.-H. Pause or no pause?: Prosodic phrase boundaries revisited. *Tsinghua Science and Technology*, Tsinghua, v. 13, n. 4, p. 500-509, ago. 2008.

VENABLES, W N; RIPLEY, B D. *Modern Applied Statistics with S*. 4. ed. New York: Springer, 2002. Available from: <<http://www.stats.ox.ac.uk/pub/MASS4>>. Access on: 10 Jan. 2018.



Automatic Segmentation of Spontaneous Speech

Segmentação automática da fala espontânea

Brigitte Bigi

Laboratoire Parole et Langage, CNRS, Aix-Marseille Université, Aix-en-Provence / France

brigitte.bigi@lpl-aix.fr

Christine Meunier

Laboratoire Parole et Langage, CNRS, Aix-Marseille Université, Aix-en-Provence / France

christine.meunier@lpl-aix.fr

Abstract: Most of the time, analyzing the phonetic entities of speech requires the alignment of the speech recording with its phonetic transcription. However, studies on automatic segmentation have predominantly been carried out on read speech or on prepared speech while spontaneous speech refers to a more informal activity, without any preparation. As a consequence, in spontaneous speech numerous phenomena occur such as hesitations, repetitions, feedback, backchannels, non-standard elisions, reduction phenomena, truncated words, and more generally, non-standard pronunciations. Events like laughter, noises and filled pauses are also very frequent in spontaneous speech. This paper aims to compare read speech and spontaneous speech in order to evaluate the impact of speech style on a speech segmentation task. This paper describes the solution implemented into the SPPAS software tool to automatically perform speech segmentation of read and spontaneous speech. This solution consists mainly in two sorts of things: supporting an Enriched Orthographic Transcription for an optimization of the grapheme-to-phoneme conversion and allowing the forced-alignment of the following events: filled pauses, laughter and noises. Actually, these events represent less than 1 % of the tokens in read speech and about 6 % in spontaneous speech. They occur in a maximum of 3 % of the Inter-Pausal Units of a read speech corpus and from 20 % up to 36 % of the Inter-Pausal Units in the spontaneous speech corpora. The UBPA measure – Unit Boundary Positioning Accuracy, of the proposed forced-alignment

system is 96.09 % accurate as regards read speech and 96.48 % for spontaneous speech with a delta range of 40 ms.

Keywords: spontaneous speech; forced-alignment; paralinguistic events.

Resumo: Na maior parte dos casos, a análise de entidades fonéticas da fala exige o alinhamento da gravação da fala com sua transcrição fonética. Entretanto, os estudos sobre segmentação automática têm sido predominantemente desenvolvidos com amostras de fala lida ou fala preparada, uma vez que a fala espontânea refere-se a uma atividade mais informal, sem qualquer preparação. Como consequência, na fala espontânea numerosos fenômenos ocorrem, tais como: hesitações, repetições, *feedback*, *backchannels*, elisões não-padrão, fenômenos de redução, palavras truncadas, e mais comumente, pronúncias não-padrão. Eventos como o riso, ruídos e pausas preenchidas também são muito comuns na fala espontânea. Este artigo objetiva comparar a fala lida e a fala espontânea a fim de avaliar o impacto do estilo de fala numa tarefa de segmentação da fala. O artigo descreve a solução implementada no programa SPPAS para a segmentação automática da fala lida e da fala espontânea. Essa solução consiste de principalmente dois aspectos: suporte para uma Transcrição Ortográfica Enriquecida para a otimização da conversão grafema-para-fonema e permissão para o alinhamento forçado (*forced-alignment*) dos seguintes eventos: pausas preenchidas, riso e ruídos. Tais eventos representam menos de 1% das ocorrências na fala lida e cerca de 6% na fala espontânea. Eles ocorrem com um máximo de 3% nas Unidades Entre-Pausas de um corpus de fala lida e de 20% a 36% nas Pausas Entre-Unidades de corpora de fala espontânea. As medidas APFU – Acurácia no Posicionamento de Fronteiras de Unidade, do sistema de alinhamento forçado (*forced-alignment system*) proposto são de 96% de acerto no que diz respeito à fala lida e 96,48% para a fala espontânea, com uma variação delta de 4 ms.

Palavras-chave: fala espontânea; sistema de alinhamento forçado (*forced alignment system*); eventos paralinguísticos

Submitted on January 9th, 2018

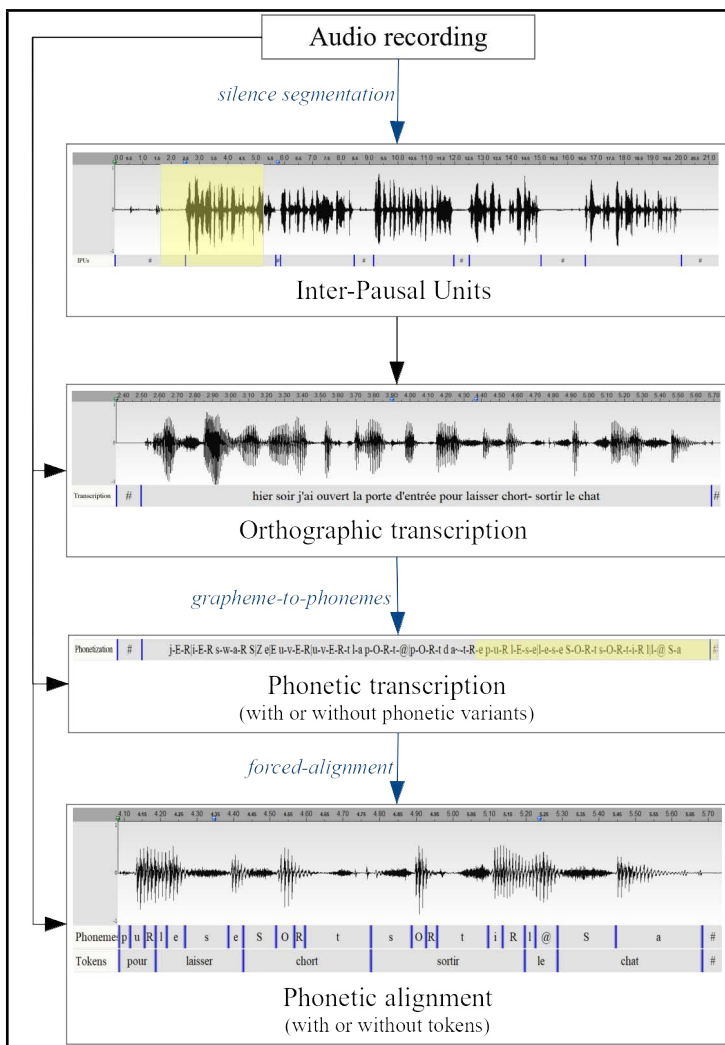
Accepted on March 23rd, 2018

1 Introduction

Speech segmentation is the process of identifying boundaries between speech units in the speech signal and determining when in time these occur. Figure 1 illustrates the full process; a blue arrow refers to a step that can be processed automatically while a black arrow refers to

a manual one. After recording the speech signal, an automatic silence segmentation algorithm creates Inter-Pausal Units (IPUs), orthographic and/or phonetic transcription is then performed, followed by the forced-alignment task, which fixes the time alignment of the sounds with the speech signal.

FIGURE 1 – Automatic speech segmentation: full process
A blue arrow indicates an automatic or semi-automatic task



Speech segmentation is important for analyzing correlations between linguistic categories such as words, syllables, or phonemes to the corresponding acoustic signal, articulatory signal, etc. In the past, phonetic studies have mostly been based on limited data. According to current trends, however, phonetic studies are expected to be built on the acoustic analysis of a large quantity of speech data and must be statistically validated. The first step in most acoustic analyses inevitably involves the time alignment of recorded speech sounds with their phonetic annotation. Segmenting and labeling of speech data have to be highly reliable. However, manual segmentation is extremely time-consuming and unlikely to be considered as a possibility if several hours of speech are to be segmented and labeled. Manual alignment has been reported to take between 11 and 30 seconds per phoneme (LEUNG; ZUE, 1984) or taking up to 400 times real time (GODFREY *et al.*, 1992). Manual alignment is too excessively time-consuming, burdensome and expensive to be commonly employed for aligning large corpora. Consequently, automatic speech segmentation is of great help for phoneticians. Knowledge of phoneme boundaries is also necessary for undertaking research on human speech processing. Moreover, research fields such as sociolinguistics and psycholinguistics depend on accurate speech transcription and segmentation at phone-level.

Determining the location of known phonemes is also important for a number of speech applications. When developing an Automatic Speech Recognition system (ASR), a robust context-dependent acoustic model is required. The latter is a statistical representation of sounds, commonly including all the phonemes of a given language and the silence. The model is trained from data sets of examples, i.e. annotated data time aligned with audios, but “good initial estimates ... are essential” when training the Gaussian Mixture Model parameters (RABINER; JUANG, 1993, p. 370). Given this context, forced-alignment is the method most commonly used in the creation of the training sets of annotated data for large speech corpora.

One of the other well-known uses of a speech segmentation system is multimedia indexing: it is necessary to provide an efficient methodology for the indexing of multimedia data for further retrieval. There is a need to index audio-video materials, and speech recognition can be used to create searchable transcripts for audio indexing in digital video libraries. Many systems have been reported in the literature; for

instance, to name but one, Moreno *et al.* (1998) proposed a recursive algorithm to perform speech segmentation for indexing long audio files. The main difference between aligning for indexing and aligning for acoustic analyses is related to precision threshold: if an offset of 2 seconds is acceptable for indexing, it is inconceivable for acoustic analysis purposes. An acceptable offset for acoustic purposes would be up to 80 ms.

Against this background, depending on the final application of the task, the system has to face different difficulties like live-audio alignment (vs batch alignment), which is done on live audio recordings and requires the aligner to manage run-time memory dynamically; like an inaccurate orthographic transcription which implies for the aligner to be designed in such a way that it can correct such erroneous points; like long audio files, which implies using strategies to manage the large amount of data; like when a high accuracy is expected for the further analyses.

The current state-of-the-art in Computational Linguistics allows many annotation tasks to be semi or fully-automated. Several toolboxes are currently available which can be used to automate the tasks (the blue ones in Figure 1). For a researcher looking for such automatic annotations, it is difficult to evaluate their usefulness and usability. Some are mainly dedicated to Computer Scientists and some are designed for Linguists. To decide about their usefulness and usability, the following have to be considered: the license, the ease of use, the type of data the tool is designed for, the supported languages or the possibility of adding a new one, and its compatibility with other annotated data. Before using any automatic annotation tool/software, it is also important to consider its error rate and how those errors can affect whatever further purpose the annotated corpora are used for. In fact, the error rate may significantly increase if the data, on which the system was trained, significantly differs from the new data to be processed. Then, another issue an automatic annotation system has to face is to consider the different types of data, particularly those related to speech style.

Shriberg (2005) has identified “four fundamental properties of spontaneous speech that present challenges for spoken language applications”: recovering hidden punctuation, coping with disfluencies, allowing for realistic turn-taking, hearing more than words. In the context of speech segmentation, the main problem among this list is to cope with disfluencies, e.g. repetitions, repairs, hesitations, etc. Shriberg

(1996) also showed that disfluencies are not ‘noise’ in speech “but rather show systematic distributions in various dimensions”. She examined filled pauses, repetitions, substitutions, insertions, deletions and speech errors, and observed that except filled pauses, they are all correlated with characteristics of the speech produced. Filled pauses however are correlated with socio-linguistic variable. Clark *et al.* (2002) re-considered the status of the English *uh* and *um*, commonly defined “filled pauses”, e.g. the audible counterparts to silent pauses, and argue that they are “words – interjections, with all the properties that this implies”. From the phonetic point of view, Shriberg (1999) examined the filled pauses, repetitions, repairs and false starts and concluded that they affect several phonetic aspects of speech. She observed changes in “segment durations, intonation, word completion, voice quality, vowel quality, and coarticulation patterns”. It mainly concerns the first two regions of the disfluency, i.e. the ones whose can be “removed to yield a fluent version of the utterance” (the reparandum and the repair). Other studies proved that the pronunciation of *the* as “*thee*” was strongly correlated with disfluent contexts, when followed by a filled pause, a pause or a repetition (TREE; CLARK, 1997). The same trend has been observed in other function words (e.g. *to* and *a*) with similar pronunciation alternations (BELL *et al.*, 2003). Most of these aspects can have an impact on speech segmentation for both the grapheme-to-phoneme and the alignment tasks.

This paper aims to highlight the differences between read speech and spontaneous speech given the specific context of the automatic speech segmentation task. It first focuses on the speech characteristics mainly related to different speech styles. Then some existing solutions to automate the speech segmentation task are presented. The paper describes several French corpora whose segmentation demand automatic speech segmentation systems particularly adapted to spontaneous speech. Finally, quantitative and qualitative results are given for the forced-alignment task.

2 Spontaneous speech

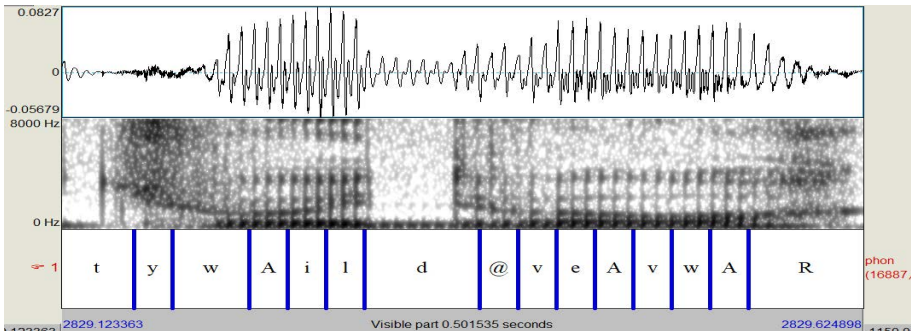
Since the end of the 20th century, studies on speech production have moved toward the analysis of more consequent corpora. Linguists used to build their own corpora, which were more generally limited in size (containing isolated words or short sentences). The apparition

of consequent corpora is due, for the most part, to the need for wider data in order to evaluate speech tools provided by automatic processing researches. Consequent corpora most generally contain more natural speech (recordings of Broadcast News, interviews, etc.) than what linguists have used for previous analysis. At the same time speech tools allowing automatic speech analyses were developed. This led linguists to analyze these consequent speech data. However, the exploitation and analyses of more casual speech raised new challenges both for linguists and for automatic speech evaluation. Indeed, the nature of casual speech is characterized by many specificities that do not appear in controlled speech.

Speech sounds produced in unconnected and prepared speech are quite easy to identify and describe. In this case, speech production is characterized by slow speech rate and speech variation is quite limited. On the other hand, speech extracted from natural and casual situations is characterized by rapid but also irregular speech rates, word truncations and phoneme reductions (JOHNSON, 2004), etc. Indeed, spontaneous speech is produced within a dynamic communicative situation. This dynamic situation involves linguistic routines and constraints, which lead to a reorganization of sound production, and then to massive variation. These characteristics result in high difficulties when speech flow has to be annotated in discrete phonetic units.

In particular, speech reduction has been of special interest since studies on spontaneous speech have become more common. It has been shown that the amount of reduction in spontaneous speech is greater than expected (JOHNSON, 2004). Different speaking styles may provide various amounts of reduction phenomena, depending on the degree of control in speech. A significant difficulty for automatic alignment tools is that reduction is not systematic to one phoneme discrete deletion. Indeed, several studies (ADDA-DECKER *et al.*, 2008; MEUNIER, 2013) have shown that, quite often, phoneme reduction results in phoneme coalescence (several phonemes are merged into one segment, Figure 2). These instances are quite frequent and are generally not perceived by transcribers. Consequently, perceived phonemes are aligned on speech signal as discrete phonetic units (Figure 2).

FIGURE 2 – Automatic alignment of the sequence “*tu vois, il devait avoir*” (you know, he should have). The effective realization shows that several phonemes are merged within one segment.



Moreover, casual speech is characterized by several elements, which do not appear in controlled or read speech. In particular, laughter, coughing, mouth noises, etc. appear very frequently in conversation. Several studies (OGDEN, 2001) point that *clicks*, for example, are used in a linguistic way in order to structure oral discourse. These elements do not belong to phonological language inventories. However, they are present in casual speech and automatic tools have to identify them in order to provide correct phonetic alignment.

One of the problems considering spontaneous speech is that read speech and spontaneous speech show major differences (ROUAS *et al.*, 2010). Indeed, the difference between a highly controlled corpus such as a read and isolated word on the one hand, and very relaxed conversation on the other hand, is also materialized by several varieties of speech types that provide specific characteristics. Variations are also found within the same style due to conditioning factors such as: the social situation, the degree of preparation, the number of interlocutors, etc. In other words, the number of reductions, repetitions and other linguistic phenomena in speech productions may vary according to the degree of control that the situation requires.

3 Automatic speech segmentation

Speech segmentation can be divided into two task categories. In the first category, the system must deal with data where transcriptions are approximate, which means that errors and omissions in the transcription are frequent. The ALISA system, for example, is dedicated to this category

(STAN *et al.*, 2016); it can align speech with imperfect transcripts in any alphabetic language. Another example is JTrans, a system performing speech segmentation on very long audio files (CERISARA *et al.*, 2009). These systems are mainly dedicated to other automatic analyses like ASR, automatic indexation, etc. In the second category, the system requires performant and accurate orthographic or phonetic transcripts in order to produce the best alignment possible. This kind of system is mainly dedicated to linguists. This paper focuses on the second category, in order to create a system with high accuracy- or at least high enough accuracy for both read speech and spontaneous speech in further studies in Phonetics and Prosody. Segmenting at the phonetic level is required in particular for the extraction of parameters such as duration, fundamental frequency or intensity within each phoneme.

Any automatic speech segmentation system requires knowledge about the language to be recognized. Such resources should define the linguistic property of the target language: recognition unit and audio properties of each unit. Typically, a unit is a word, and the following must be available for the system to work:

- a lexicon of the target language that defines the words to be recognized;
- a word dictionary, i.e. their pronunciation as a sequence of phonemes including pronunciation variants or not;
- an acoustic model, i.e. a stochastic model of input waveform patterns per phoneme. Systems can be based on the use of various types of models, including the well-known Hidden Markov Models (HMM). Hand-transcribed speech training data are required to build a highly accurate acoustic model.

The lexicon and the word dictionary constitute the linguistic resources necessary to perform the automatic phonetic transcription task, and the acoustic model is required for the automatic phonetic alignment task.

3.1 Automatic phonetic transcription

In the initial stage, the automatic system converts to the given orthography into a sequence of phonemes; this task is named “grapheme-

to-phoneme” in Figure 1. It implies two sub-tasks for the system. Firstly, the given orthographic transcription is normalized into units. Secondly, the units are converted into a sequence of phonemes with or without pronunciation variants. This phonetization can be performed either by a set of pronunciation rules or can be based on a pronunciation dictionary. The availability of these systems to support spontaneous speech implies coping with all the speech phenomena described in section 2. For example, the phonetization system must include a solution for generating the pronunciation of missing words like broken words, regional words, mispronunciations, it has to be able to deal with pronunciation variants, and in general with any kind of disfluency.

To deal with speech variability, the system can add alternative expected phonetic segments so that it lets the automatic alignment choose the best option. This grapheme-to-phoneme conversion assumes that it can generate a result that contains the correct pronunciation. However, casual speech is highly variable. Numerous studies have investigated the automation of pronunciation variations. Statistical decision trees to generate alternate word pronunciations were used in (RILEY *et al.*, 1999). A phonetic-feature-based prediction model is presented in (BATES *et al.*, 2007). Recently, (LIVESCU *et al.*, 2016) proposed an “approach of modeling pronunciation variation in terms of the time course of multiple sub-phonetic features”.

In previous works (BIGI, 2014, 2016), we proposed a multilingual text normalization system and a multilingual phonetization system. The methods are designed to be as language-and-task-independent as possible: this makes it possible to add new languages with a significant time-reduction compared to the entire development of such tools. The approach is also relevant to the present study because it functions indifferently with any kind of speech style. The system supports an Enriched Orthographic Transcription (EOT), which allows the transcriber to include the following:

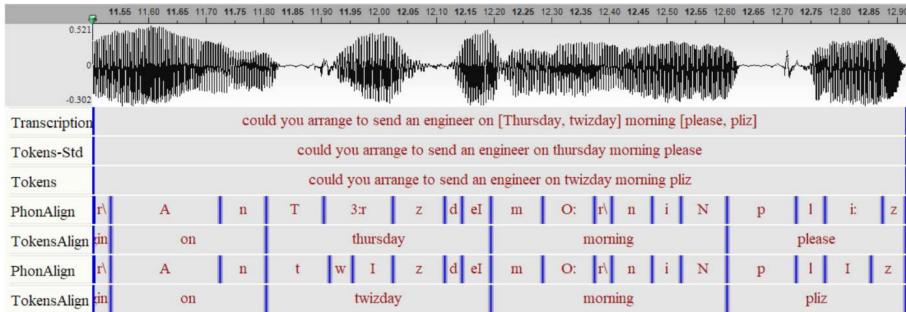
- a broken word is noted with a ‘-’ at the end of the token string;
- a noise is noted ‘*’; it can be a breath, a cough or an unintelligible segment, ...
- laughter is noted by a ‘@’;
- a short pause is noted by a ‘+’;

- an elision is mentioned between parenthesis, like thi(s);
- an unexpected pronunciation is noted with brackets like this [example, eczap];
- a comment of the transcriber is noted with braces or brackets like {this comment} or [this other comment];
- an unexpected liaison is surrounded by ‘=’;
- a morphological variant is noted like this <ice scream, I scream>;
- a proper noun may be surrounded by ‘\$’ symbols like \$ Alan Turing \$;
- regular punctuation and character case are accepted.

The system does not require all these phenomena to be mentioned in order to work; nevertheless, this specific convention makes it possible to annotate all perceivable disfluencies. The user can thus integrate the degree of enrichment he requires into the transcription.

When these speech phenomena are mentioned in the manual orthographic transcription, it significantly increases the result of the grapheme-to-phoneme conversion (BIGI *et al.*, 2012), either by using a rule-based system or a dictionary-based system. On the basis of a standard orthographic transcription, the dictionary-based system results in 10.8 % errors on read speech up to 14.5 % on conversational speech. By using the proposed enrichments of the orthographic transcription, errors were significantly reduced to 8.2 % on read speech and 9.5 % on conversational speech. So this multilingual approach of automatic phonetization performs well and very accurately for different types of speech. Furthermore, the EOT associated with the appropriate automatic systems can help in tackling the problems of the grapheme-to-phoneme conversion on various types of data. For example, Figure 3 illustrates the use of the enrichment to transcribe (tiers *Transcription*), normalize (tiers *Tokens*, *Tokens-Std*), phonetize and time-align (tiers *PhonAlign*, *TokensAlign*) a Spanish native speaker while reading an English text. The automatic text normalization, phonetization and alignment were performed firstly on the standard version and secondly on the modified one, both automatically extracted from the EOT.

FIGURE 3 – Example of enriched orthographic transcription and automatic speech segmentation on read speech by a learner speaker



The approach based on EOT improves the accuracy of the speech segmentation result for the grapheme-to-phoneme conversion and consequently for the forced-alignment; and it opens research opportunities for Linguists. However, the enrichment of the transcription is time consuming for the user. One way to speed up the process is to add the most frequent reductions into the pronunciation dictionary. For example, in French, the word “*parce que*” p-a-R-s-k (*because*) is often pronounced p-s-k, or a pronoun like “*tu*” t-y (*you*) is pronounced t. But adding them into a pronunciation dictionary supposes that such frequent reductions were previously identified.

Frequent reductions can be detected automatically as proposed in (SCHUPPLER *et al.*, 2008): a lexicon of canonical phonemic representations of the words was used in a first stage and a second experiment was carried out with a lexicon that had been enriched with pronunciation variants. “These variants were generated by applying reduction rules to the canonical transcriptions of the words”, thanks to a forced-alignment system. Alternatively (or additionally), the EOT can be a way to identify and to add the frequent pronunciation variants to the dictionary manually (MEUNIER, 2012).

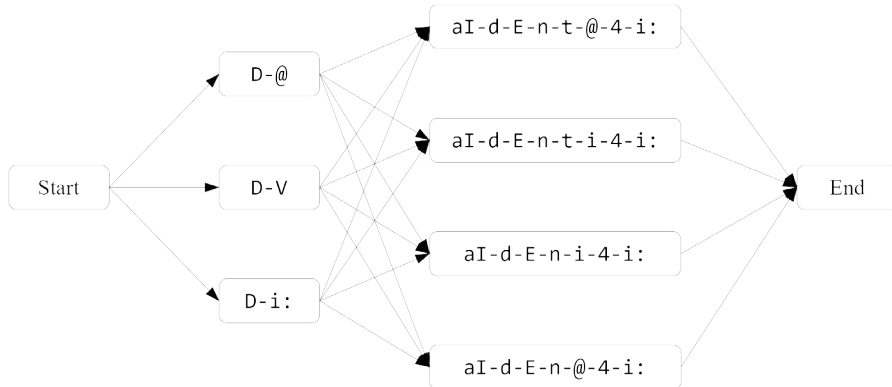
Finally, a reasonable level of orthographic enrichment has to be determined. On the one hand, as it has been said, enrichment of transcription is time-consuming work. On the other hand, human cannot hear very fine variations or reductions so that many segmental variations and reductions cannot be transcribed. As a result, orthographic transcription should be enriched by uncommon variations easy to identify by with a simple transcription code; and the automatic speech

segmentation system has to identify the most common variations, for instance, “*tu vois*” (*you see*) often pronounced t-y-w-a, with /v/ deleted.

3.2 Automatic phonetic alignment

While a phonetic transcription is available, a forced-alignment system has to be implemented in order to estimate where the sounds are located in the acoustic signal. For that specific purpose, Automatic Speech Recognition engines are useful. Any ASR system can perform automatic forced-alignment. The task is named “forced-alignment” because the phoneme alignment is obtained by forcing the ASR result to be the given phonetic sequence. A language model or a grammar has to be given to the ASR system to fix constraints of the sequence of phonemes, it can eventually include pronunciation variants, i.e. several possible paths to the ASR result like the example of Figure 4.

FIGURE 4 – Example of representation of a grammar for the forced-alignment task: the English sequence “the identity”. The word “the” can be phonetized into 3 different sequences of phonemes and the word “identity” into 4.



There are several cross-platform toolkits for building a recognition system. Notable among these are: HTK – Hidden Markov Model Toolkit (YOUNG; YOUNG, 1993), CMU Sphinx (LAMERE *et al.*, 2003), Open-Source Large Vocabulary CSR Engine Julius (LEE *et al.*, 2001), RASR (RYBACH *et al.*, 2009) and Kaldi (POVEY *et al.*, 2011). Among this list, HTK, RASR and Kaldi have to be compiled to prepare for their intended use. These systems are open-sources but HTK and RASR require users

to register and the HTK license precludes distribution or sub-licensing to third parties. These systems are distributed as toolboxes and can be used only by means of the command-line interface; they require knowledge and skill about speech processing.

In order to facilitate their use, a large number of wrappers have been developed, although, they all require an aligner to be previously installed. They make it possible to automatically time-align speech in an easier way than the direct use of the ASR system. Additionally, some of them include several features like training an acoustic model, estimating statistical distributions of annotated data or performing requests for data, etc. Table 1 reports the main characteristics of these existing wrappers. It includes Train&Align (BROGNAUX *et al.*, 2012), P2FA (YUAN; LIBERMAN, 2008), Prosodylab-Aligner (GORMAN *et al.*, 2011), The Munich Automatic Segmentation System MAUS web service (KISLER *et al.*, 2017), PraatAlign (LUBBERS; TORREIRA, 2016), and SPPAS (BIGI, 2015). This table does not present a fully comprehensive list and is restricted to the freely available tools whose developers can be contacted. It has to be noted that most of the systems use a specific representation of phonemes, except SPPAS in which phonemes are represented in X-SAMPA and a plugin makes conversion into the International Phonetic Alphabet possible.

The first column of Table 1 is the name of the wrapper. The interface column mentions the way the system communicates with users: CLI for a Command-line User Interface, GUI for a Graphical User Interface and Web for a web service. The third column refers to the list of languages the system is able to deal with: it means that acoustic models are included in the wrapper. The quality of such models correlates strongly with the data used for the training and it is possible that it doesn't match with the new data to be processed. Some of the systems are not provided with acoustic models and/or propose to train the model directly from the data to be aligned; but this supposes that there is enough of such data. The fourth column indicates what the ASR system the wrapper is based on. The next column lists the operating systems, except in the case of a web service. The last column indicates the list of file formats the wrapping system is able to cope with, without distinguishing inputs/outputs for reasons of clarity.

It should be noted that with the same acoustic model and the same aligner, the wrappers should produce the same phoneme alignment result.

For example, we would expect WebMAUS to produce the same results as PraatAlign because the acoustic models included in the former were picked up by the latter. Depending on its licensing condition, it is then feasible for an acoustic model of a wrapper to be included in and used by another one: it is then not scientifically relevant to directly compare alignment results of wrappers.

TABLE 1 – Some alignment wrappers freely available on the Internet

Wrapper name	Interface	Language	Aligner	Operating system	File format
Train&Align	Web	-	HTK	-	txt
P2FA	CLI	eng	HTK	Windows Linux MacOS	json, textgrid
Prosodylab-Aligner	CLI	eng	HTK	Windows Linux MacOS	txt, textgrid
WebMAUS	Web	28+	HTK	-	txt, textgrid, par, xml, csv
PraatAlign	GUI	spa, nld	HTK	Windows Linux MacOS	textgrid
SPPAS	CLI GUI	eng, fra, spa, ita, cat, pol, yue, jpn, nan, pcm, cmn, (kor), (por)	Julius HTK	Windows MacOS Linux	txt, textgrid, trs, eaf, tdf, lab, antx, csv, ctm, stm, sub, srt, anvil, mrk, xra

Finally, anyone who has automatic alignment to perform can easily access these systems and choose the most appropriate one depending on his/her own needs: the interface, the supported language, the aligner that has to be previously installed, the input/output file format, etc. All of these forced-alignment systems are capable of achieving acceptable results on the alignment of read speech.

However, despite the availability of numerous systems, the alignment of spontaneous speech remains a challenging task: previous

work to improve the accuracy of the phoneme boundaries for spontaneous speech is sparse. Among the above-mentioned systems, our system (SPPAS) is the only one to bring forward a full solution to this issue: The next sections state the reasoning behind the development of this solution and describe its implementation and accuracy.

4 Collected corpora

In order to compare automatic speech segmentation of read speech and spontaneous speech, we conducted an inventory then a selection of some existing data. We picked out French data so that they constitute as far as possible two homogeneous sets: read and spontaneous speech.

All the selected corpora were recorded in an attenuated-sound booth with one microphone per speaker. Each audio signal was automatically segmented into IPU's - Inter-Pausal Units that are segments of speech surrounded by silent pauses over 200 ms, and time-aligned on the speech signal. The IPU's boundaries were all manually checked. For each of the speakers an orthographic transliteration was then provided. The transcription process followed the specific convention described in section 3. However, the extent of the enrichment depends on the corpus, but in all corpora, the following are mentioned: filled pauses, laughter, noises, disfluencies (repetitions, broken words, etc), unusual pronunciations and short pauses. The main difference between the enrichment of the transcription concerns the amount of elisions. Finally, for this study, we normalized all the corpora with the same version of our text normalization system, and we phonetized with our phonetizer (see section 3). We then expected to achieve the best possible uniformity of the data: the only thing that differed was the speech style.

Table 2 summarizes the corpora that were gathered for the present study. The first column indicates the name of the corpus. The second column refers to the manual transcription available, i.e. one or several between:

- a. both phonetized and time-aligned;
- b. a standard orthographic transcription;
- c. an enriched orthographic transcription.

It was also expected that these transcriptions be double-checked. Unfortunately, this was not the case for *AixMapTask*. The third column indicates the duration of spoken segments, excluding the duration of silences; and below is the number of speakers. The last column indicates the speech style. For some corpora, only a part of the corpus was extracted to ensure that all the above-mentioned criteria were respected.

TABLE 2 – Description of the corpora

Corpus name	Transcription	Speech duration Nb speakers	Speech style
Data collected locally (audio)	Phonetized, Time-aligned	2 min 2 spks	Reading (words/sentences)
<i>Europe</i> (audio)	Phonetized, Time-aligned, Enriched ortho.	33 min 6 spks	Political debate (radio broadcast)
<i>Eurom1</i> (audio)	Standard ortho.	28 min 10 spks	Reading (5 paragraphs)
<i>AixOx</i> (audio)	Enriched ortho.	110 min 10 spks	Reading (10 paragraphs)
<i>Typaloc</i> (audio)	Enriched ortho.	32 min 19 spks	Reading (2 texts)
<i>Typaloc</i> (audio)	Enriched ortho.	39 min 4 spks	Conversation (interview)
<i>AixMapTask</i> (audio-video)	Enriched ortho.	163 min 10 spks	Conversation (task-oriented)
<i>CID</i> (audio-video)	Enriched ortho.	7h30min 16 spks	Conversation (casual dialog)
<i>Cheese</i> (audio-video)	Enriched ortho.	63 min 8 spks	Reading a joke; Conversation (casual dialog)

Europe corpus (PORTES, 2004) is a debate recorded from a radio broadcast. It involves two journalists and four invited speakers debating on the European Union and particularly on its frontiers.

A part of the French *Eurom1* corpus was extracted. It consists in “40 passages made of five thematically linked sentences, showing a coherent semantic structure so as to induce a correct prosodic structure at each sentence level” (CHAN *et al.*, 1995).

AixOx (HERMENT *et al.*, 2014) replicates *Eurom1* with a larger number of speakers and texts to read: 40 paragraphs are read by 10 speakers.

TYPALOC (MEUNIER *et al.*, 2016) is composed by several corpora of reading (words and texts) and spontaneous speech (interviews) produced by healthy speakers and by patients affected by different types of dysarthria. The healthy speakers selected for this study read two short texts and had a free discussion (8-17 min) with an experimenter who invited them to tell some stories from their own life.

The audio-visual condition of *Aix Map Task* is a corpus of audio and video recordings of task-oriented dialogues (GORISH *et al.*, 2014). The experimental design follows the standard rules of Map Task experiments: participants were allowed to say anything necessary to accomplish their communicative goals. In this face-to-face condition, the two participants could see each other.

Corpus of Interactional Data - CID (BERTRAND *et al.*, 2008) is an audio-video recording of 8 dialogs involving two participants, 1 hour of recording per session. One of the following two topics of conversation was suggested to participants: conflicts in their professional environment or funny situations in which participants may have found themselves.

Cheese (PRIEGO-VALVERDE; BIGI, 2016) is also an audio-video recording of dialogs involving two participants. They had received the task to read each other a canned joke chosen by the experimenters, and then to converse as freely as they wished to for the rest of the interaction. Figure 5 illustrates the recording conditions of the audio-visual corpora.

FIGURE 5 – Experimental condition of audio-visual corpora: *CID*, *AixMapTask*, *Cheese*



5 Corpora distributions

5.1 Distribution of tokens

Tables 3 and 4 indicate the number of tokens of each corpus for read speech and spontaneous speech respectively. Any speech production is considered as a token: a word, an interjection, a feedback, etc. The tables also indicate the amount of some events, limited to the 3 following categories:

1. The filled pause. In French, the filled pause has a standard spelling (“*eah*”); it is then uniformly transcribed in corpora and easy to identify automatically.
2. Laughter. They are manually indicated in the orthographic transcription by the ‘@’ symbol.
3. Other events are all named under the generic term “noise”. They can be breathing in or out, coughs, or any kind of noise in the microphone that is produced by the speaker. They are manually indicated in the orthographic transcription by the ‘*’ symbol. The recording of such noises depends highly on the quality and the position of the microphone. Thus, drawing conclusions on the differences between noises in the corpora should be avoided.

TABLE 3 – Tokens and paralinguistic events in read speech

Corpus	Number of tokens	% of filled pause	% of laughter	% of noise
AixOx	28,408	0.014 %	0 %	0.134 %
Eurom1	6,912	0 %	0.014 %	0 %
Cheese (read part only)	1,086	0.092 %	0.829 %	0.184 %
Typaloc (read part only)	6,377	0 %	0.047 %	0.047 %

TABLE 4 – Tokens and events in spontaneous speech

Corpus	Number of tokens	% of filled pause	% of laughter	% of noise
Europe	7,566	6.014 %	0.013 %	0.264 %
Typaloc	7,534	2.933 %	0.186 %	1.434 %
AixMapTask	37,979	2.285 %	0.635 %	2.607 %
CID	126,260	3.997 %	1.221 %	0.870 %
Cheese	16,829	2.793 %	2.246 %	0.434 %

These tables obviously highlight the fact that the selected events are much less frequent in read speech than in spontaneous speech. The majority of such events in read speech (less than 1 %) concerns laughter in *Cheese*, probably because the speakers were reading a joke. Table 4 shows that the amount of events is differently distributed according to the data. All the corpora of spontaneous speech contain a high percentage of filled pauses, ranging from 2.3 % up to 6 %. Actually, the *Europe* corpus contains a significantly higher amount of filled pauses than the others, which is not surprising for a political debate on the radio; and for the same reason, this debate contains only one example of laughter. On the contrary, the casual conversations contain more laughter. The interviews and the map-task contain a more reasonable amount of laughter probably because during the recording both interviewees or participants have to complete a task.

The distribution of tokens through the overall corpora shows a surprising regularity. Indeed, when selecting the ten most frequent words in the corpora, the four function words “*de*” (*of*), “*la*” (*the*), “*et*” (*and*), “*le*” (*the*) are present, except in *Cheese*. These four words are highly frequent in spontaneous speech as well as in read speech. This suggests that they are essential in order to structure and construct oral speech. Other words appear frequently according to the characteristics of each corpus. For example, “*est*” (*is*) is systematically present in the

inventory of frequent words in spontaneous corpora, but is absent from read corpora one. The feedback marker “*ouais*” (*yeah*) is also ranked at the 4th or 5th position in *MapTaskAix*, *CID* and *Cheese*.

In all the spontaneous corpora, noise and laughter are not in the ten most frequent tokens, except laughter, which is at the 5th position in *Cheese*. The filled pause is included in the five most frequent tokens; and in all but *Europe* and *CID*, it is the most frequent token.

5.2 Filled pause, laughter and noise events

In the context of speech segmentation, the challenge of many events is not so much their grapheme-to-phoneme conversion but lies rather in their time-alignment on the acoustic signal. It is important then to determine where these events are located relative to speech. The first column of Table 5 indicates how the percentages of times such events are surrounded by silences, i.e. they are the unique token of the IPU. In this situation, the automatic forced-alignment is not involved because segmentation had already been accomplished at the first stage of the process (by the IPU's segmentation task). All 3 of the other columns are related to a situation in which they have to be segmented by the forced-alignment system:

1. when the event starts with an IPU, the alignment system has to fix the boundary between the event and the next sound;
2. when the event ends with an IPU, the alignment system has to fix the boundary between the last sound of the IPU and the event;
3. when the event is inside an IPU, i.e. the paralinguistic event is surrounded by speech and/or another event so that the alignment system has to fix the starting and ending boundaries of the event.

Table 5 clearly indicates that the filled pauses occur close to speech, 98.53 % of their items start are inside or end an IPU. To a lesser extent, we observe that laughter items and noises are also close to speech. This table clearly highlights the need for the automatic speech segmentation system able to handle these events.

TABLE 5 – Percentage of the events depending on their left and right context

	surrounded by silences	starting an IPU	ending an IPU	inside an IPU
filled pause	1.47 %	11.80 %	28.96 %	57.77 %
laughter	34.72 %	19.10 %	29.05 %	17.13 %
noise	20.86 %	28.03 %	11.63 %	39.48 %

Moreover, the forced-alignment task performs an optimization algorithm on the whole IPU so that a misalignment of a sound necessarily has consequences on the closest sounds or even further. Table 6 indicates the amount of IPUs of the corpus and the percentage of these IPUs that are concerned by the selected events. In read speech, they are observed in a maximum of 3.32 % of the IPUs (*Cheese* corpus). However, 20 % up to 36 % of the IPUs include at least one of the events we have identified.

TABLE 6 – Amount of IPUs in which the events are occurring

Corpus	# total IPUs	IPUS with filled pause	IPUs with laughter	IPUs with noise	IPUs with any event
AixOx (read)	2,724	0.15 %	0	1.28 %	1.40 %
Cheese (read)	241	0.41 %	3.32 %	0.83 %	3.32 %
Europe (spont)	875	35.88 %	0.11 %	2.29 %	35.89 %
Typaloc (spont)	522	28.25 %	2.68 %	14.94 %	35.82 %
AixMapTask (spont)	6,126	12.16 %	3.67 %	13.52 %	20.60 %
CID (spont)	13,631	27.32 %	10.25 %	7.52 %	32.14 %
Cheese (spont)	2,675	14.62 %	12.45 %	2.73 %	21.16 %

The following IPUs were extracted from *Typaloc* and *Cheese* spontaneous corpora. They clearly illustrate the phenomena quantified in Table 5. They also illustrate that the events often co-occur in an IPU

like shown in Table 6. Indeed, compared to read speech, spontaneous speech is characterized by sequences of speech which include frequent paralinguistic events. More precisely, we observe that the presence of these events is related to the type of spontaneous speech: laughter is quite infrequent in interviews or guided tasks; conversely, it is more frequent in conversations (*CID, Cheese*). Moreover, filled pauses are relatively frequent in most IPUs within spontaneous corpora.

Example 1 from *Typaloc (spont)*:

*donc euh des choses euh genre euh canard à l'orange des choses
comme ça qui demandent euh une préparation un peu plus subtile
une surveillance*

*(then uh things uh like uh duck in orange sauce something like that
which require uh a slightly more subtle preparation a supervision)*

Example 2 from *Cheese (spont)*:

tu vas avec ton père euh il repart avec mille chameaux à @

*(you travel with your father uh he goes back home with one
thousand camels @)*

6 Forced-alignment: read vs. spontaneous speech

The previous section highlights the fact that some events are so frequent that a forced-alignment system should be able to automatically time-align them, particularly in case of spontaneous speech whatever the context (interview, conversation, etc.). This section reports on the possibility for an acoustic model to include a model for each of these events. It measures its relevance. We aimed at developing an automatic alignment system that could place boundaries with accuracy comparable in both speech styles: read and spontaneous speech.

6.1 Test corpus and evaluation method

A test corpus was manually phonetized and segmented by one expert, then revised by another one. The data files of the test set were randomly extracted from the training set and removed from the latter. It includes two subsets:

- read speech: 127 seconds of *AixOx* (1776 labels);
- conversational speech: 141 seconds of *CID* (1833 labels).

The read speech test set includes 4 speakers, reading 44 IPUs for which 9 contain noise items; and the spontaneous speech test set includes 12 speakers, with 27 IPUs for which 20 contain the selected events. Table 7 presents the detailed distribution of the labels in both data sets. For the read speech, the noise represents 0.58 % of the labels; and for spontaneous speech the 3 events represent 1.80 % altogether of the labels to be aligned. The system includes the following 31 phonemes:

- vowels: A/ E e 2 i O/ 9 u y
- nasalized vowels: a~ U~/ o~
- plosives: p t k b d g
- fricatives: f v s z S Z
- consonant nasals: m n
- liquids: l R
- glides: H j w

where A/ represents a or A, O/ represents o or O and U~/ represents e~ or 9~, in SAMPA code.¹

Most of the boundaries between phonemes were easy to fix manually in the spectrograms with a precise position in time due to clear differences in intensity or voicing. But speech is a continuous process and dividing it into discrete, non-overlapping, and directly consecutive units necessarily involves ambiguities and discrepancies. So, no particular segmentation can be claimed to be the correct one. Among others, it was observed in (HOSOM, 2008) that the agreement between two expert humans is, on average, 93.78 % within 20 ms on a variety of English corpora.

¹ French SAMPA proposed by J. C. Wells at: <<http://www.phon.ucl.ac.uk/home/sampa/french.htm>>.

TABLE 7 – Labels of the test subsets

Label	Read speech	Conversational speech
phoneme	1736	1791
filled pause	0	24
laughter	0	5
noise	10	4
short pause	30	9

For the experiments, we estimated the Unit Boundary Position Accuracy (UBPA) that has been widely used in previous studies. It measures what percentage of the automatic-alignment boundaries are within a given time threshold of the manually aligned boundaries. UBPA is an automatic evaluation of the place of boundaries that measures the deviation between the corresponding segment boundaries placed by humans and the system. This kind of error analysis reports a quantitative information that allows knowing the overall performances of the systems.

6.2 Forced-alignment without and with selected events

For the acoustic models, all the labels are 5-state HMMs. Typically, the HMM states are modeled by Gaussian mixture densities. Models were trained from 16 bits, 16,000 Hz sample-rated wav files. The Mel-frequency cepstrum coefficients (MFCC) along with their first and second derivatives were extracted from the speech in a common way, 25 coefficients altogether: Delta coefficients appended (*_D*); Absolute log energy suppressed (*_N*); Cepstral mean subtracted (*_Z*); Cepstral C0 coefficient appended (*_0*).

Two series of acoustic models were created; a series depends on the amount of speech that was used to train. The training of the first series has no particular influence on the filled pause, laughter and noise events; therefore, it can be considered a state-of-the-art system. In the second series, the acoustic models include specific models for them.

For the first series, the acoustic models were created from the read-speech training set only. The models for the filled pause, noise and laughter were set to the prototype model. This prototype results of the

HCompv command of the HTK toolkit. See (BIGI, 2014) for details about the training procedure that we implemented into the *acmtrain.py* script and *acm* package of SPPAS.

For the second series, the acoustic models of the first series were modified, in order to focus on evaluating the impact of the use of the three events. Specific models were trained for all of them from the spontaneous speech data. During this training procedure, filled pause items were phonetized *fp*, noises *gb* and laugh items *lg*. The latter models were introduced in the previously created acoustic models, replacing the already existing ones.

Therefore, the only difference between the first and the second series of acoustic models lies in the models of the three selected events. We then measured the impact of adding models for these events in the acoustic model of the system for both read speech and spontaneous speech. Figures 6 and 7 display the UBPA of two such series of acoustic models. Each series of models was separately evaluated on the read-speech (Figure 6) and on the spontaneous-speech (Figure 7) test sets. All models were initialized with the same two minutes of manually phonetized and time-aligned data. The X-axis represents the amount of read speech data that was added during the training stage, represented in seconds, among the three corpora for which an enriched orthographic transcription is available: *Typaloc*, *AixOx* and *Cheese*. These models were then trained from the two minutes manually time-aligned plus randomly picked-up files in these read-speech corpora. Five runs were performed for each amount of data, and the displayed accuracy is the average of their UBPA. A final model was trained with all available read-speech data representing about 3h of *Typaloc*, *AixOx*, *Cheese* and *Eurom1* altogether.

FIGURE 6 – UBPA (in percentage, with a delta of 20 ms) of acoustic models on read speech

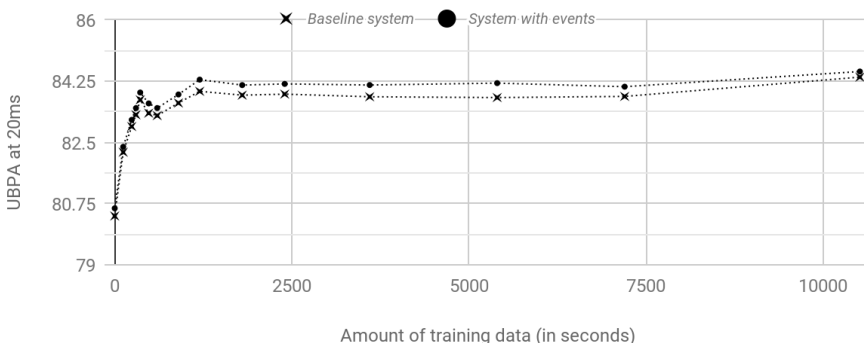
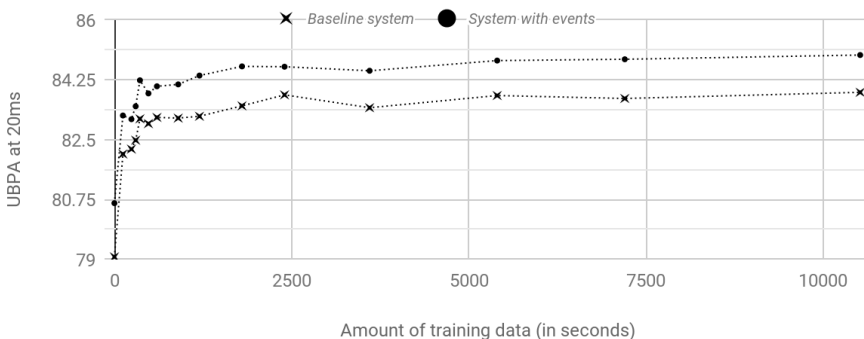


FIGURE 7 – UBPA (in percentage, with a delta of 20 ms) of acoustic models on spontaneous speech



Both figures show that the initial model, trained from the 2 minutes of manually time-aligned data is already quite good (about 80 % accuracy) and so it constitutes a good initialization model for further training. We observe that from 2 to 20 minutes of training material, the accuracy increases significantly in all conditions. Then the models reach a relatively stable state, i. e. a slow but steady increase with small time-to-time variations. These results enable advice to be given to data producers who are expecting automatic speech segmentation on a given language: at least 2 minutes of manually time-aligned data and at least 20 minutes of properly transcribed data have to be created to form the acoustic model.

More importantly, these figures highlight important differences between the accuracy of the models with or without the three selected events on spontaneous speech. As was expected, the differences in read speech are not truly significant, due to the absence of filled pauses and laughter in the test corpus. The significant improvements on spontaneous speech reflected what we described in the previous section: these events are very frequent and the forced-alignment system has to be adapted. The accuracy of the model trained with all data increases from 83.88 % to 84.97 % on spontaneous speech that represents 1.09 % absolute gain and 6.76 % relative gain. The UBPA of the same model on read speech is 84.53 %.

Finally, we noticed that the UBPA at 40 ms of the models trained with all read-speech training data reaches 95.64 % on read speech and 95.67 % on spontaneous speech when the events are introduced. Experiments of this section made it possible to conclude that forced-alignment can reach very close performances on read speech and on spontaneous speech as soon as the acoustic model includes the 3 selected events: filled pause, laughter and noise.

6.3 Relative importance of the selected events

Our system is not the only one to deal with these events. For example, P2FA includes a model for laughter and three different models for noises. This section aims at comparing the relative impact of the events and constructing a final acoustic model able to cope optimally with the most varied speech styles.

In the scope of obtaining the best acoustic model, a new model has been created by adding the manually phonetized and time-aligned *Europe* corpus to the training data. The latter is made of all of the read-speech corpora. The filled pause (fp), noise (gb) and laughter (lg) were then added to the acoustic model as in the experiments described in the previous section. It should be noted that adding the spontaneous corpora described in Table 2 drastically decreases the accuracy of the model. So, these latter data were used only to train the models of lg, fp and gb but not to train the models of the phonemes.

Table 8 presents the accuracies of this final model at various delta values. Adding *Europe* corpus in the training procedure significantly increases the accuracy of the model on both spontaneous speech and read

speech. This final model reaches a good overall alignment performance whatever the speech style and so the system has the ability to withstand variations in speech.

TABLE 8 – UBPA (%) of the final acoustic model depending on the delta value (Europe data were included in the training set)

	20 ms	30 ms	40 ms	50 ms	80 ms
read speech	85.54	93.75	96.09	97.82	99.22
spont. speech	86.10	93.94	96.48	97.62	99.19

Table 9 quantifies the impact of each event on the alignment of spontaneous speech. It shows that the use of a trained-noise model instead of the prototype does not really affect accuracy. With only 4 occurrences in the test set, it is not surprising but it could have slightly done. However, it should be noted that on read speech, the UBPA at 40 ms of the model without *gb* is 95.92 % and it increases to 96.09 % with *gb*. This result brings us to conclude that the use of a generic model for all noises does not have very much impact on the accuracy. However, even if the test set contains only 5 laughter items, creating a specific model impacts significantly on the results: the accuracy at 40 ms grows from 96.05 % without *lg* to 96.48 % with *lg*, which represents an absolute gain of 0.43 % and relative gain of 10.89 %. Finally, the most important event that has to be represented in an acoustic model is the filled pause. In the previous section, we observed that filled pauses represent 2.28 % to 6.01 % of the tokens in the corpora of spontaneous data. In the test set, 24 items have to be time-aligned, over the 1833 labels; *fp* then represents 1.31 % of the labels to be aligned. Table 9 shows that at 40 ms, the accuracy of the model without *fp* is 94.81 % and Table 8 shows that the final acoustic model with a trained *fp* model is 96.48 %. The absolute gain is therefore 1.67 % and the relative gain is 32.18 %.

TABLE 9 – UBPA (%) on the spontaneous-speech test set of the acoustic model depending on the event

	20 ms	30 ms	40 ms	50 ms	80 ms
model without gb	86.26	94.00	96.43	97.56	99.13
model without lg	85.83	93.62	96.05	97.13	98.54
model without fp	84.96	92.43	94.81	96.00	97.72

To complete this analysis, we should mention that, with exception of our system, all systems that support French language use sound 2 to represent the filled pause instead of using the prototype as we tested in our previous experiments. We then evaluated the accuracy of our model when the model of **fp** is substituted by the model of the vowel 2. UBPA at 40 ms is 95.67 % and at 80 ms is 98.53 %. It results in a significantly better accuracy compared to the use of the prototype (line 3 of Table 9), but a specific model achieves better accuracy (line 2 of Table 8). It can, thus, be concluded that the use of a vowel that is acoustically close to the filled pause is a good alternative in cases where no data is available to train a specific model for the filled pause but the latter is the preferable solution.

6.4 Analysis of the major errors

The previous experiments were based on the use of the UBPA. This accuracy measure allows us to detect what are called fine errors, “when the automatic segment boundary is not 100 % overlapping the corresponding manually placed segment boundary” (KVALE, 1994). UBPA has proved its effectiveness in comparing the performance of models; however, it does not highlight relevant information about the nature, extent, and timing of errors. A qualitative error analysis allowed us to estimate whether the deviations from human annotation introduce any bias.

We examined the errors when the automatic boundary is 80 ms over the manual one. This occurs 15 times in the read speech test corpus and 16 times in the spontaneous test set.

On read speech, it is noticeable that the errors are uniformly distributed over the files of the test. Five of the shifted boundaries lie between a phoneme and a short pause and one between a short pause

and a phoneme: this means that 20 % of the short pauses are not properly time aligned. This highlights a weakness in our model that we will have to investigate in future works. Other errors are sparse.

Contrarily to read speech, on spontaneous speech, errors are grouped into five IPU of four different speakers. Figure 8 reports on the most salient errors concentrated in an IPU for the sequence of tokens “na na na na na na”. The speaker just wanted to report an undescribed discourse and he produced a hypo-articulated sequence. If the transcription is compatible to the production of the speaker, the automatic aligner failed in finding correct boundaries because of phoneme coalescence. During this sequence of speech, 6 errors were referenced. The system firstly missed the second token “na” by setting too long a duration of the first A/. The last 4 phonemes of this sequence are following the principle of a forced-alignment system: they are “forced” even if the system can’t find them in the signal and the minimum duration is assigned (30 ms) to each of them. Figure 9 illustrates another typical case of errors in cascades. The system fails to find the beginning of the laughter and assigns the phoneme A/ to the first “sound” of the laughter - which is acoustically close to a A. This error has an impact on the segmentation of the sequence of 4 phonemes: k-t-w-A/.

FIGURE 8 – Misalignment on the spontaneous data set in the sequence of speech “na na na na na na”

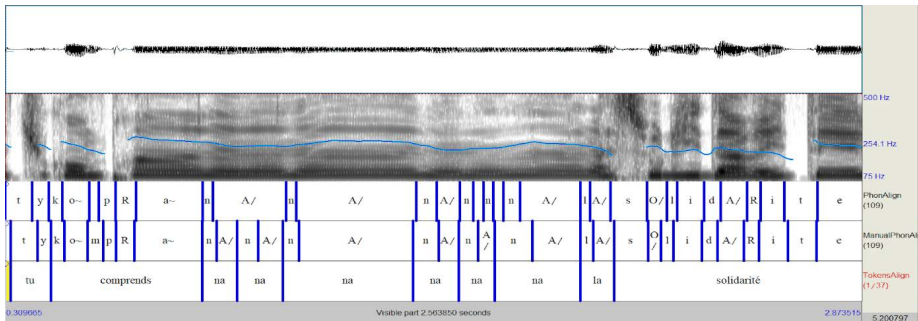
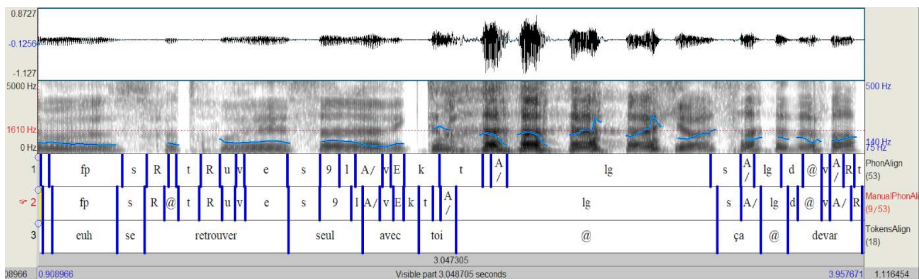


FIGURE 9 – Misalignment on the spontaneous data set with a laughter



6.5 Analysis of the segmentations

We propose detailed quantitative analyses of the differences between the manual and the automatic annotation for each phoneme in terms of 3 variables:

- the duration;
- the position of the boundary for the beginning;
- the position of the boundary for the end.

These comparisons are plotted by means of an R script, wrapped in the script we developed and included into SPPAS software tool. It evaluates the accuracy of an acoustic model with a more specific view. These diagrams provide precious information to the Linguists for a better understanding of the results from the automatic system.

Figures 10 and 11 represent this kind of result. In order not to overload this document, both figures show the duration of the phonemes only (automatic vs. manual). A positive value in the duration graph means that the duration of the phoneme is higher in the automatic segmentation than in the manual one. On read speech, we can observe that it mainly concerns g, h, w and z. The observation of the two other graphs indicates that in both cases, the start boundary is slightly earlier and the end boundary is slightly later than expected. On the contrary, a negative value in the duration graph means that the duration of the phoneme of the automatic segmentation is smaller than the manual one. This is significantly the case for the consonants p and v because the start position of the automatic system is generally later than expected but the

end boundary is close to the expected one. U~/ is also reduced by the automatic system because of an anticipated end boundary. The most significant reduced phoneme on read speech is 2 for which the start boundary is later than expected and the end is earlier. On the contrary, the automatic system correctly aligns 2 in spontaneous speech. We can also observe that the alignment of the filled pause is as good as the alignment of any phoneme with a perfect average duration and a very reasonable variation in the range of 20 ms; the whiskers are not very far either. However, durations of noise are systematically over-rated by 20 ms on average contrarily to the duration of laughter, which is underestimated by 20 ms on average.

From a global view of these figures, for the vowels the differences between read speech and spontaneous speech mainly concern 2 and 9; and for consonants the system performs alignment significantly differently for the phonemes, p, t, k and H.

FIGURE 10 – Differences between the manual and the automatic annotation on read speech

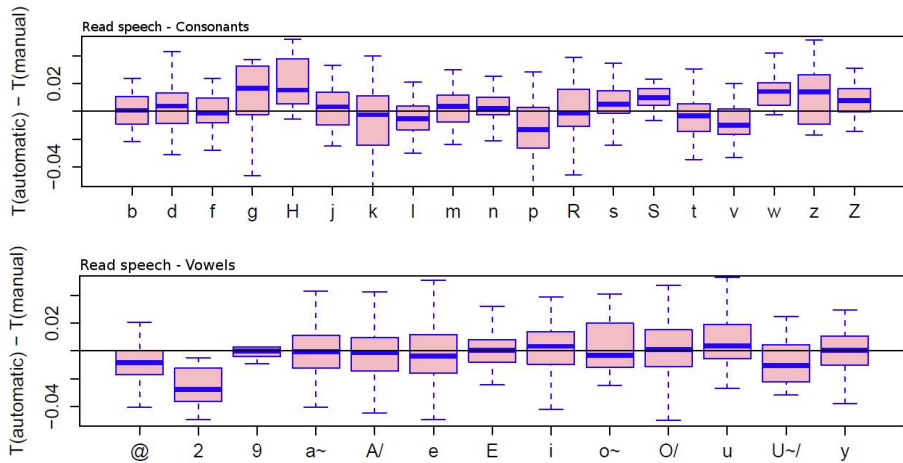
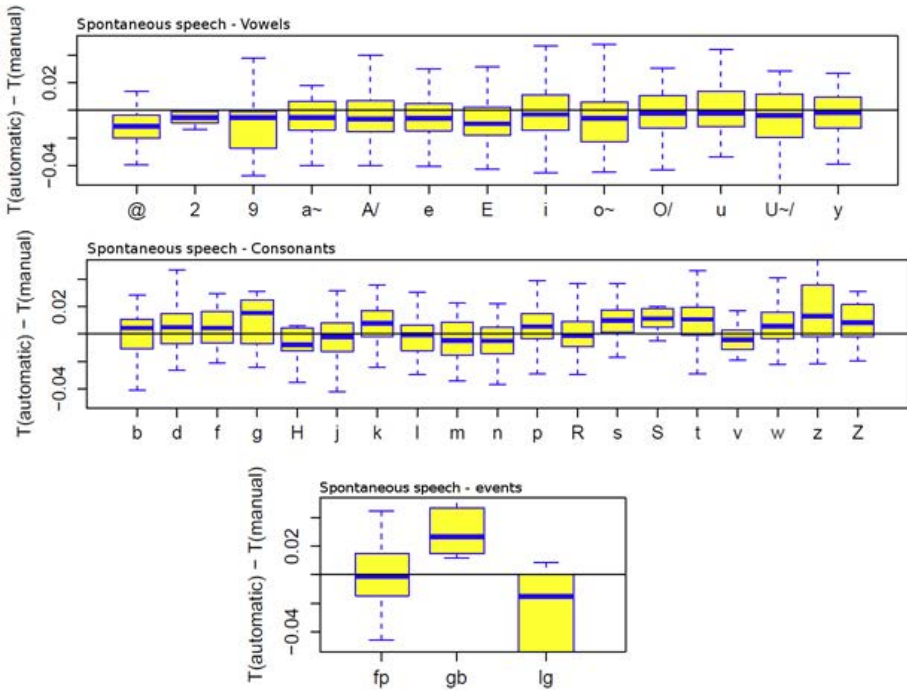


FIGURE 11 – Differences between the manual and the automatic annotation on spontaneous speech



6.6 The aligner

All the forced-alignment results mentioned in the previous sections were estimated by using a wrapper of the Julius CSR engine, version 4.2.2. Finally, we assessed the impact of the aligner on the accuracy of the forced-alignment task. We estimated the results if we use a wrapper of the *HVite* command of HTK, version 3.4.1. In this experiment, only the aligner has changed; we used the acoustic model described in section 6.3 that obtains results as in Table 8 when Julius is used.

Table 10 indicates the UBPA by using a system based on the *HVite* command. The second line indicates the difference of accuracy with the system based on *Julius*.

TABLE 10 – UBPA of the final acoustic model with the *HVite* aligner. The second line indicates if the accuracy with *HVite* is lesser, higher or equal than with *Julius*.

	20 ms	30 ms	40 ms	50 ms	80 ms
read speech	84.59 (-0.95)	94.03 (+0.28)	96.15 (+0.06)	97.82 (=)	99.33 (+0.11)
spont. speech	83.34 (-2.76)	92.64 (-1.30)	96.27 (-0.21)	97.62 (=)	99.13 (-0.06)

Compared to Table 8, Table 10 clearly shows that *Julius* performs better than *HVite* on spontaneous speech particularly when the delta of the UBPA is less than 50 ms. On read speech, results are either lesser, higher or equals with *Julius* or with *HVite* depending on the precision of the accuracy. Then, the aligner system has an impact on the alignments mainly for fine errors, and it has a relatively bigger impact on spontaneous speech than on read speech. Future work will have to investigate on the other aligner systems, including Sphinx, Kaldi and RASR.

Conclusion

This paper addressed the problem of automatic-speech segmentation for both read speech and spontaneous speech. Compared to read speech, spontaneous speech differs in two major issues: 1/ a significant increase of speech variations, and 2/ the embedding, *within speech*, of events such as laughter, coughing, etc. These two differences have to be considered by automatic systems because they have an impact on phonetic-acoustic analyses and because their study is relevant for linguistic and conversation analysis. In the system we propose, most of the difficulties involving the first point are tackled by the grapheme-to-phoneme system: broken words, repetitions, elisions, mispronunciations, etc. We briefly presented a full solution for the grapheme-to-phoneme conversion and introduced the EOT - Enriched Orthographic Transcription. This solution was designed to be as language- and-task independent as possible. Based on a relevant orthographic transcription and a pronunciation dictionary, the system can work on speech of any language and of any style, including disfluencies. This paper attracted more attention on the second point about embedded events and on the forced-alignment task. The phoneme alignment of read

speech can actually be done quite easily thanks to state-of-the-art systems freely available on the web. However, the automatic forced-alignment of spontaneous speech remain a challenge.

The distributions of 3 selected events in several corpora were presented: the filled pause, laughter and noise. We quantified these events in both read speech and various styles of spontaneous speech. They were observed in a maximum of 3.32 % of the IPU's in a read speech corpus while in spontaneous speech 20 % up to 36 % of the IPU's include at least one of these 3 events. Experiments were performed to estimate their impact on the forced-alignment task. They led us to conclude that forced-alignment can reach very close performances on read speech and on spontaneous speech as soon as the acoustic model includes the events. This result implies that the acoustic model is robust enough to cope with speech reductions and variations, even on spontaneous speech. Qualitative and quantitative analyses of the results pointed a slight weakness of our model for the alignment of short pauses. However, we observed a very close quality in the alignment of phonemes between read speech and spontaneous speech. The alignment of the filled pause performs also as well as the alignment of any phoneme; durations of noise events are overrated by 20 ms on average contrarily to the duration of laughter, which is underestimated, by 20 ms on average.

In the context of this study, we created a robust acoustic model for French language. This model will be included in version 1.9.5 of SPPAS and distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License. The file is saved in HTK-ASCII format² in order to allow the model of each sound to be extracted and re-used in another acoustic model, as soon as the latter is based on the same acoustic parameters. Moreover, the Python library and the scripts to train an acoustic model or to estimate the UBPA will also be included in the software under the terms of the GNU Public License version 3. Both will be available as a functionality in the CLI.

² This format is neither compressed nor encoded. It is simply a readable file that can be easily edited with any text editor.

Authors' contributions

Brigitte Bigi designed the study, performed the experiments and implementation of the research. Christine Meunier brought information about the problematic of the segmentation of spontaneous speech and contributed to the interpretation of the results. Both wrote the manuscript.

References

ADDA-DECKER, M.; GENDROT, C.; NGUYEN, N. Contributions du traitement automatique de la parole à l'étude des voyelles orales du français. *Traitement Automatique des Langues – ATALA*, [s.l.], v. 49, n. 3, p. 13-46, 2008.

BATES, R. A.; OSTENDORF, M.; WRIGHT, R. A. Symbolic phonetic features for modeling of pronunciation variation. *Speech Communication*, Elsevier, v. 49, n. 2, p. 83-97, 2007. Doi: <https://doi.org/10.1016/j.specom.2006.10.007>

BELL, A.; JURAFSKY, D.; FOSLER-LUSSIER, E.; GIRAND, C.; GREGORY, M.; GILDEA, D. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America*, [s.l.], v. 113, n. 2, p. 1001-1024, 2003. Doi: <https://doi.org/10.1121/1.1534836>

BERTRAND, R.; BLACHE, P.; ESPESSER, R.; FERRÉ, G.; MEUNIER, C.; PRIEGO-VALVERDE, B.; RAUZY, S. Le CID – Corpus of Interactional Data – Annotation et Exploitation Multimodale de Parole Conversationnelle. *Traitement Automatique des Langues*, – ATALA, [s.l.], v. 49, n. 3, 2008.

BIGI, B. The SPPAS participation to Evalita 2014. In: ITALIAN CONFERENCE ON COMPUTATIONAL LINGUISTICS CLiC-it, 1; INTERNATIONAL WORKSHOP EVALITA, 4., 2014, Pisa, Italy. *Proceedings...* Pisa: Pisa University Press, 2014. v. 2, p. 127-130.

BIGI, B. A Multilingual Text Normalization Approach. In: VETULANI, Z.; MARIANI, J. (Ed.). *Human Language Technology Challenges for Computer Science and Linguistics, LTC 2011*. Lecture Notes in Computer Science. Berlin: Springer Berlin Heidelberg, 2014. v. 8387, p. 515-526. Doi: https://doi.org/10.1007/978-3-319-14120-6_42

BIGI, B. SPPAS – Multi-lingual Approaches to the Automatic Annotation of Speech. *The Phonetician*, International Society of Phonetic Sciences, v. 111-112, p. 54-69, 2015.

BIGI, B. A phonetization approach for the forced-alignment task in SPPAS. In: VETULANI, Z.; USZKOREIT, H.; KUBIS, M. (Ed.). *Human Language Technology Challenges for Computer Science and Linguistics, LTC 2013*. Lecture Notes in Computer Science. Berlin: Springer Berlin Heidelberg, 2016. v. 9561, p. 397-410. Doi: https://doi.org/10.1007/978-3-319-43808-5_30

BIGI, B.; BERTRAND, R.; PÉRI, P. Orthographic Transcription: which enrichment is required for phonetization? In : INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 8., 2012, Istanbul, Turkey. *Proceedings...* Istanbul: European Language Resources Association, 2012. p. 1756-1763.

BROGNAUX, S.; ROEKHAUT, S.; DRUGMAN, T. *et al.* Train&Align: A New Online Tool for Automatic Phonetic Alignment. In: *IEEE Spoken Language Technology Workshop*, 4., 2012, Miami, EUA. *Proceedings...* Miami: [s.n.], 2012. p. 416-421. Doi: <https://doi.org/10.1109/SLT.2012.6424260>

CHAN, D.; FOURCIN, A.; GIBBON, D.; GRANSTROM, B.; HUCKVALE, M.; KOKKINAKIS, G.; KVALE, K.; LAMEL, L.; LINDBERG, B.; MORENO, A.; MOUROPOULOS, J.; SENIA, F.; TRANCOSO, I.; VELD, C.; ZEILIGER, J. “EUROM- A Spoken Language Resource for the EU”. In: EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND SPEECH TECHNOLOGY, 4., 1995, Madrid, Spain. *Proceedings...* Madrid: [s.n.], 1995. v. 1, p. 867-870.

CLARK, H. H.; TREE, J. E. F. Using *uh* and *um* in spontaneous speaking. *Cognition*, Elsevier, v. 84, n. 1, p. 73-111, 2002. Doi: [https://doi.org/10.1016/S0010-0277\(02\)00017-3](https://doi.org/10.1016/S0010-0277(02)00017-3)

CERISARA, C.; MELLA, O.; FOHR, D. JTrans, an open-source software for semi-automatic text-to-speech alignment. In: ANNUAL CONFERENCE OF THE INTERNATIONAL SPEECH COMMUNICATION ASSOCIATION, 10., 2009, Brighton, United Kingdom. *Proceedings...* Brighton: International Speech Communication Association, 2009. p. 1823-1826.

- GODFREY, J. J.; HOLLIMAN, E. C.; McDANIEL, J. SWITCHBOARD: Telephone speech corpus for research and development. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, 1992, San Francisco, USA. *Proceedings...* San Francisco: IEEE, 1992. p. 517-520.
- GORISCH, J.; ASTÉSANO, C.; GURMAN BARD, E.; BIGI, B.; PRÉVOT, L. Aix Map Task corpus: The French multimodal corpus of task-oriented dialogue. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 9., 2014, Reykjavik, Iceland. *Proceedings...* Reykjavik: [s.n.], 2014. p. 2648-2652.
- GORMAN, K.; HOWELL, J.; WAGNER, M. Prosodylab-Aligner: A Tool for Forced Alignment of Laboratory Speech. *Canadian Acoustics*, Canada, v. 39, n. 3, p. 192-193, 2011.
- HERMENT, S.; TORTEL, A.; BIGI, B.; HIRST, D. J.; LOUKINA, A. AixOx, a multi-layered learners' corpus: automatic annotation. Specialisation and Variation in Language Corpora. *Linguistic Insights: Studies in Language and Communication*, Oxford, v. 179, p. 41-76, 2014.
- HOSOM, J. P. Speaker-independent phoneme alignment using transition-dependent states. *Speech Communication*, Elsevier, v. 51, n. 4, p. 352-368, 2008. Doi: <https://doi.org/10.1016/j.specom.2008.11.003>
- JOHNSON, K. Massive Reduction in Conversational American English. In: YONEYAMA, K.; MAEKAWA, K. (Ed.). *Spontaneous Speech: Data and Analysis*. Proceedings of the 1st Session of the 10th International Symposium. Tokyo, Japan: The International Institute for Japanese Language, 2004. p. 29-54.
- KISLER, T.; REICHEL, U. D.; SCHIEL, F. Multilingual processing of speech via web services. *Computer Speech & Language*, Elsevier, v. 45, p. 326-347, 2017. Doi: <https://doi.org/10.1016/j.csl.2017.01.005>
- KVALE, K. On the Connection Between Manual Segmentation Conventions and “errors” Made by Automatic Segmentation. In: INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING, 3., 1994, Yokohama, Japan. *Proceedings...* Yokohama: Acoustical Society of Japan, 1994. p. 1667-1670.
- LAMERE, P.; KWOK, P.; GOUVEA, E.; RAJ, B.; SINGH, R.; WALKER, W.; WARMUTH, M.; WOLF, P. The CMU SPHINX-4 speech recognition

system. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, 2003, Hong Kong. Hong Kong: IEEE, 2003. v. 1. Doi: 10.1109/ICASSP.2003.1202277

LEE, A.; KAWAHARA, T.; SHIKANO, K. Julius – an open source real-time large vocabulary recognition engine. In: EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY, 7., 2001, Aalborg, Denmark. *Proceedings...* Aalborg: [s.n.], 2001. p. 1691-1694.

LEUNG, H. C.; ZUE, V. W. A. Procedure for Automatic Alignment of Phonetic Transcriptions with Continuous Speech. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, 1984, San Diego, USA. *Proceedings...* San Diego: IEEE, 1984. v. 9, p. 73-76. Doi: <https://doi.org/10.1109/ICASSP.1984.1172426>

LIVESCU, K.; JYOTHI, P.; FOSLER-LUSSIER, E. Articulatory feature-based pronunciation modeling. *Computer Speech & Language*, Elsevier, v. 36, p. 212-232, 2016. Doi: <https://doi.org/10.1016/j.csl.2015.07.003>

LUBBERS, M.; TORREIRA, F. PraatAlign: an interactive Praat plug-in for performing phonetic forced alignment. 2016. Available at: <<https://github.com/dopefishh/praatalign>>. Retrieved on : 05/28/2018.

MEUNIER, C. Contexte et nature des réalisations phonétiques en parole conversationnelle. In : JOURNEES D'ETUDE SUR LA PAROLE, 2012, Grenoble, France. Actes... Grenoble : AFCP ; ATALA, 2012. p.1–8.

MEUNIER, C. Phoneme deletion and fusion in conversational speech. In: EXPERIMENTAL APPROACHES TO PERCEPTION AND PRODUCTION OF LANGUAGE VARIATION, 2013, Copenhagen, Denmark. *Proceedings...* Copenhagen: University of Copenhagen, 2013.

MEUNIER, C.; FOUGERON, C.; FREDOUILLE, C.; BIGI, B.; CREVIER-BUCHMAN, L. *et al.* The TYPALOC Corpus: A Collection of Various Dysarthric Speech Recordings in Read and Spontaneous Styles. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION CONFERENCE, 10., 2016, Portorož, Slovenia. *Proceedings...* Portorož: ELRA, 2016. p. 4658-4665.

MORENO, P. J.; JOERG, C.; THONG, J-M. V. *et al.* A recursive algorithm for the forced alignment of very long audio segments. In:

INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING, 5., 1998, Sydney, Australia. *Proceedings...* Sydney: ISCA Archives, 1998. http://www.isca-speech.org/archive/icslp_1998

OGDEN, R. Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of the International Phonetic Association*, Cambridge, v. 31, n. 1, p. 139-152, 2001. Doi: <https://doi.org/10.1017/S0025100301001116>

PORTES, C. *Prosody and Discourse: phonetic specificity, discursive ecology and pragmatic meaning of the “implication contour”*. 2004. Thesis (PhD) – Université de Provence - Aix-Marseille I, 2004.

POVEY, D., GHOSHAL, A., BOULIANNE, Gilles, *et al.* The Kaldi speech recognition toolkit. In: IEEE WORKSHOP ON AUTOMATIC SPEECH RECOGNITION AND UNDERSTANDING, 2011, Waikoloa, Hawaii. *Proceedings...* Waikoloa: IEEE Signal Processing Society, 2011.

PRIEGO-VALVERDE, B.; BIGI, B. Smiling behavior in humorous and non humorous conversations: a preliminary cross-cultural comparison between American English and French. In: INTERNATIONAL SOCIETY FOR HUMOR STUDIES CONFERENCE, 2016, Dublin, Ireland. Oral Presentation. Available at: <<https://hal.archives-ouvertes.fr/hal-01455222>>. Retrieved on : 05/28/2018.

RABINER, L. R.; JUANG, B. H. *Fundamentals of Speech Recognition*. Englewood Cliffs: PTR Prentice Hall, 1993. v. 14.

RILEY, M.; BYRNE, W.; FINKE, M. *et al.* Stochastic pronunciation modelling from hand-labelled phonetic corpora. *Speech Communication*, Elsevier, v. 29, n. 2, p. 209-224, 1999. Doi: [https://doi.org/10.1016/S0167-6393\(99\)00037-0](https://doi.org/10.1016/S0167-6393(99)00037-0)

ROUAS, J-L.; BEPPU, M.; ADDA-DECKER, M. Comparison of spectral properties of read, prepared and casual speech in French. In: LANGUAGE RESOURCE AND EVALUATION CONFERENCE, 7., Malta, 2010. *Proceedings...* Malta: University of Malta, 2010. p. 606-611.

RYBACH, D.; GOLLAN, C.; HEIGOLD, G.; HOFFMEISTER, B.; LÖÖF, J.; SCHLÜTER, R.; NEY, H. The RWTH Aachen University Open Source Speech Recognition System. In: ANNUAL CONFERENCE OF THE INTERNATIONAL SPEECH COMMUNICATION ASSOCIATION,

10., Brighton, U.K., 2009. *Proceedings of the Interspeech 2009...*, Brighton: ISCA Archive, 2009. p. 2111-2114.

SCHUPPLER, B.; ERNESTUS, M.; SCHARENBERG, O.; BOVES, L. Preparing a corpus of Dutch spontaneous dialogues for automatic phonetic analysis. ANNUAL CONFERENCE OF THE INTERNATIONAL SPEECH COMMUNICATION ASSOCIATION, 9., Brisbane, Austrália, 2008. *Proceedings...* Brisbane: ISCA Archive, 2008. p. 1638-1641.

SHRIBERG, E. Disfluencies in switchboard. In: INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING, 4., 1996, Philadelphia, EUA. *Proceedings...* Philadelphia: ISCA Archives, 1996. p. 11-14.

SHRIBERG, E. Phonetic consequences of speech disfluency. In: INTERNATIONAL CONGRESS ON PHONETIC SCIENCES, 14., San Francisco, EUA, 1999. *Proceedings...* San Francisco: University of California, 1999. p. 619-622.

SHRIBERG, E. Spontaneous speech: How people really talk and why engineers should care. In: EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY, 9., Lisbon, Portugal, 2005. *Proceedings...* Lisbon: ISCA, 2005.

STAN, A.; MAMIYA, Y.; YAMAGISHI, J.; BELL, P.; WATTS, O.; CLARK, R. A.; KING, S. ALISA: An automatic lightly supervised speech segmentation and alignment tool. *Computer Speech & Language*, Elsevier, v. 35, p. 116-133, 2016. Doi: <https://doi.org/10.1016/j.csl.2015.06.006>

TREE, J. E. F.; CLARK, H. H. Pronouncing “the” as “thee” to signal problems in speaking. *Cognition*, Elsevier, v. 62, n. 2, p. 151-167, 1997. Doi: [https://doi.org/10.1016/S0010-0277\(96\)00781-0](https://doi.org/10.1016/S0010-0277(96)00781-0)

YOUNG, S.J.; YOUNG, S.J. *The HTK hidden Markov model toolkit: Design and philosophy*. Cambridge: University of Cambridge, Department of Engineering, 1993.

YUAN, J.; LIBERMAN, M. Speaker identification on the SCOTUS corpus. *Journal of the Acoustical Society of America*, [s.l.], v. 123, n. 5, 2008. Doi: <https://doi.org/10.1121/1.2935783>



Acoustic Correlates of Prosodic Boundaries in French: A Review of Corpus Data

Correlatos acústicos de fronteiras prosódicas em francês: uma revisão de dados de *corpora*

George Christodoulides

Language Sciences and Metrology Unit, Université de Mons, Mons / Belgium

george@mycontent.gr

Abstract: In this article we investigate the acoustic correlates of prosodic boundaries in French speech. We compare the prosodic structure annotation performed by experts in two multi-genre corpora (Rhapsodie and LOCAS-F). A uniform analysis procedure is applied to both corpora. The results show that the main acoustic correlates of prosodic boundaries are silent pauses and pre-boundary syllable lengthening. Pitch movements contribute to the perception of boundaries but are essentially correlates of boundary function, rather than boundary strength. Two levels of four-level annotation of boundary strength in the Rhapsodie corpus (periods and packages) correspond to the two-levels of strength in the LOCAS-F corpus.

Keywords: prosody; speech segmentation; prosodic boundaries; corpus linguistics; French.

Resumo: Neste artigo investigamos os correlatos acústicos de fronteiras prosódicas da fala em língua francesa. Comparamos a anotação da estrutura prosódica efetuada por anotadores *experts* em dois corpora multigêneros (Rhapsodie e LOCAS-F). Um procedimento de análise uniforme é aplicado a ambos os *corpora*. Os resultados indicam que os principais correlatos acústicos de fronteiras prosódicas são pausa silenciosa e alongamento da sílaba pré-fronteira. Movimentos de *pitch* contribuem para a percepção de fronteiras mas são essencialmente correlatos de funções de fronteira, e não de força de fronteira. Dois dos níveis de anotação dos quatro níveis de anotação de força de fronteira do *corpus* Rhapsodie (períodos e pacotes) correspondem aos dois níveis de intensidade do *corpus* LOCAS-F.

Palavras-chave: prosódia; segmentação da fala; fronteiras prosódicas; linguística de *corpus*; francês.

Submitted on May 14th, 2018Accepted on July 18th, 2018

1 Introduction

The segmentation of speech into meaningful units is central to discourse comprehension. In this respect, prosody is used by the speaker to guide the listener in reconstructing the intended segmentation and understand the message. For this reason, numerous studies have been dedicated to understanding how prosodic cues are used to signal the segmentation of an utterance, and the relationship between the prosodic segmentation and other levels of linguistic analysis, such as the syntactical structure and the information structure of speech.

Researchers working on French have particularly focused on the relationship between prosodic structure and syntactic structure. Two projects have resulted in two spoken French corpora including multiple speakers in multiple communicative situations (speaking styles), with very similar research objectives: the Rhapsodie corpus (LACHERET *et al.*, 2014) and the LOCAS-F corpus (MARTIN *et al.*, 2014). An analysis of the properties of the prosodic boundaries annotated by experts in the LOCAS-F corpus has already been presented in Christodoulides and Simon (2015); the relevant aspects of this study are repeated here for the reader's convenience.

In this article we will compare the annotation of prosodic structure in the Rhapsodie and the LOCAS-F corpora. These annotations were performed independently, by different experts in French prosody, and following different theoretical frameworks. In this study, we are interested in calculating the acoustic correlates of prosodic boundaries based on each of the two annotations and searching for similarities and differences. Our work has both a theoretical motivation and a practical application: in order to develop software tools for the automatic annotation of prosodic structure (e.g. MITTMANN; BARBOSA, 2016), appropriately-sized, publicly-available corpora are essential. We are therefore interested in exploring whether machine learning models trained on each of these two French corpora will be consistent with each other.

2 Related work

The prosodic segmentation of an utterance, as expressed by the prosodic boundary cues, is central to discourse comprehension (for a review, see CUTLER, 1997 and FÉRY, 2017). It has been shown that prosodic boundaries facilitate comprehension, by indicating the intended segmentation to the listener (e.g. SWERTS, 1997; CLIFTON *et al.*, 2002; WATSON; GIBSON, 2005; FRAZIER *et al.*, 2006). Stress, prominence and prosodic boundaries play a central role in defining the prosodic structure and arriving at a phonological description of any language (MERTENS, 2014). However, the factors contributing to the perception of prosodic segmentation are not completely understood. Phonological theories differ in the number of prosodic segmentation levels, and consequently on the number of prosodic boundary strengths. Consequently, there is no consensus on a universally-accepted, objective method of segmentation of utterances into prosodic units. Corpus resources with prosodic annotation have been compiled over the past years, including: for English, the AixMARSEC corpus (AURAN *et al.*, 2004) and the Boston University Radio News Corpus (OSTENDORF *et al.*, 1995); for French, Italian, Portuguese and Spanish, the C-ORAL-ROM collection of corpora (CRESTI; MONEGLIA, 2005); and the Spoken Dutch Corpus (SCHUURMAN *et al.*, 2003).

Although most models on French prosody admit at least three degrees of prosodic boundaries and a hierarchy of three levels of units (JUN; FOUGERON, 2000; MERTENS, 1993; ROSSI, 1999; DI CRISTO, 1999), most large-scale corpus annotations are limited to one or two degrees (e.g. the C-ORAL-ROM corpus). Furthermore, there is evidence that listeners perceive prosodic boundaries as a gradual phenomenon and in relative terms, i.e. they perceive a boundary as stronger or as weaker than the previous one.

As discussed in detail in Wagner *et al.* (2015), research on prosodic prominence can be grouped into three main perspectives: a functional, a physical and a cognitive perspective. A similar categorisation can be applied to research on prosodic segmentation, given that a prosodic boundary will render the syllable (or, more generally speaking, the right edge of a larger prosodic unit) on which it is realised prominent, in the sense that this syllable (or right edge) will stand out from its environment by virtue of its prosodic characteristics.

A functional perspective on prosodic prominence and segmentation focuses on its communicative and core linguistic functions; this approach lends itself to a categorical classification: a syllable is prominent or not, a prosodic boundary is present or not. This is the approach taken in phonological theories, that discretise the perception of boundaries and use a small number of prosodic boundary strengths (e.g. major, intermediate, minor) to define a hierarchy of prosodic units. A physical perspective will treat prosodic prominence and prosodic segmentation as a continuous rather than a categorical phenomenon, similar to a psycho-acoustic scale. Under this approach, perceptual experiments help in identifying a number of signal-related correlates to the perception of prominence or segmentation; these correlates are continuous physical quantities (e.g. duration, fundamental frequency, voice source features etc) that combine (e.g. using a linear combination formula) to give a “degree” or “score” of perceived prominence or boundary strength. A cognitive perspective focuses on perceptual processing, i.e. the way in which these phenomena are interpreted and contribute in higher-level cognitive processes. These processes are shaped by linguistic knowledge and situation-specific expectations. The cognitive perspective relies on both the functional perspective and the physical perspective. Wagner *et al.* (2015) argue that these perspectives are complementary, that they are “different parts of the same elephant”.

In the present study, we investigate the acoustic correlates of prosodic boundaries on the basis of annotations on two corpora. We are therefore taking an intermediate route between a physical perspective and a functional (and, to some extent, a cognitive) perspective. The expert annotators of these corpora were all native speakers of French and indicated the presence of prosodic boundaries based on their perception, influenced by the speech signal, their linguistic knowledge and their top-down expectations, and working within a specific functional model that determined the number of prosodic boundary strengths used in the annotation (four-level vs. two-level).

Previous research (e.g. MO *et al.*, 2008; MO, 2008; WAGNER; WATSON, 2010) suggests that silent pauses, duration, f_0 movement and phonation type are the most salient cues to prosodic boundaries. Those cues are known to be language-specific to some extent. In French, since the primary (final) accent is located on the last syllable of a prosodic unit, it co-occurs with the prosodic boundaries (cf. DI

CRISTO, 2011). However, this does not mean that French listeners cannot distinguish between prominence and prosodic phrasing, as shown in perception experiments by Astésano *et al.* (2012). Experiments with naïve listeners have identified silent pause duration, syllable duration and pitch movements as relevant acoustic correlates of prosodic prominence and prosodic segmentation in French (e.g. PORTES, 2002; SMITH, 2011).

3 Method

3.1 Corpora

The Rhapsodie corpus is a corpus covering multiple speaking styles and was created with the objective of studying the relationship between prosodic phrasing and syntax in French. The corpus samples were mainly collected from existing French corpora, including the PFC corpus (DURAND *et al.*, 2009), C-PROM (AVANZI *et al.*, 2010) and CFPP (BRANCA-ROSOFF *et al.*, 2012). The corpus contains 57 short samples (the average sample duration is 5 minutes) for a total of 3 hours of speech and 33,000 tokens. The corpus samples were balanced across four dimensions: the degree of speech planning, the degree of interactivity, the communication channel, and the main discourse strategy used by the primary speaker (oratory, argumentative, descriptive, or procedural); the corpus contains both monologues and dialogues.

In the Rhapsodie corpus, the syntactic annotation is articulated in two levels, called “micro-syntactic” and “macro-syntactic” by the authors; the main theoretical framework posits the use of “pile structures” to represent the syntactic relations of short segments of continuous speech, including self-corrections and other types of disfluencies. The prosodic annotation includes: prosodically prominent syllables annotated by experts based on their perception, using two levels (weak and strong); an annotation of disfluencies at the syllable level. (e.g. lengthening); and a prosodic structure annotation composed of intonational periods, intermediate packages, rhythmic groups and metrical feet. A perceptual boundary annotation was abandoned by the project due to poor inter-annotator agreement (LACHERET *et al.*, 2014). The four-level annotation was performed within the Autosegmental-Metrical theoretical framework.

The LOCAS-F corpus is similarly a corpus covering multiple speaking styles, including both monologues and dialogues, and was also created in order to study the relationship between prosody and syntax in French. The corpus contains 48 samples organised in 14 different speaking styles; its duration is 3.5 hours and it contains approximately 43.000 tokens. Samples from the C-PROM corpus were reused in the LOCAS-F corpus; the reused samples are 3 radio news broadcasts, 3 political public addresses, 3 scientific conference presentations, 2 radio interviews and 3 monologue narrations of life events; 75% of the C-PROM corpus is included in LOCAS-F, and C-PROM samples make up 25% of the LOCAS-F corpus.

In the LOCAS-F corpus, the syntactic annotation is articulated in two levels: a sequential, non-overlapping grouping of tokens into “functional sequences” that are further grouped into dependency clauses (a clause consisting of its root and all its dependent elements). The prosodic annotation was performed by two expert annotators. Each word was marked as being followed by a strong PB (///), an intermediate PB (//), or as not followed by any boundary (0). The annotators used the code “hesi” to indicate that they perceive the speaker was hesitating: this includes filled pauses (e.g. “euh”) and drawls. A function was also attributed to each PB, based on the shape of the corresponding intonation contour. Four types of contours were used: C (continuation), T (final prosody), S (suspense) and F (focus). This annotation was primarily based on the annotators’ perception; however, they did have visual access to the pitch contour as displayed in *Praat* (BOERSMA; WEENINK, 2017). In cases of disagreement, the annotators listened to the relevant section once again and agreed on the final prosodic boundary and contour label. Note that a “focus” contour is related to the fact that the annotator perceived an element of the utterance as being made salient, and not necessarily on a definition of prosodic prominence.

The Rhapsodie corpus is available under a Creative Commons license and can be downloaded from the project’s website (www.projet-rhapsodie.fr). The LOCAS-F corpus is not publicly available; our analyses are based on the version of the corpus that was made available to us for the study presented in Christodoulides and Simon (2015) and our subsequent work on the corpus.

3.2 Data analysis

Both corpora were imported into *Praaline* (CHRISTODOULIDES, 2014) for processing and to render the annotations comparable. The TextGrids and XML files of the Rhapsodie corpus are publicly available on the project's website; the LOCAS-F corpus is already stored as a *Praaline* SQL database but it is not yet publicly available.

We enhanced the available annotations in the corpora by applying *DisMo* multi-level annotator (CHRISTODOULIDES *et al.*, 2014) and the *Prosogram* series of scripts for intonation stylisation (MERTENS, 2004). An automated script was used to extract all potential prosodic boundary sites, i.e. all syllables at the right boundary of a multi-word unit (as annotated by *DisMo*). The script calculates multiple prosodic measures on each syllable, including:

- the duration of a subsequent silent pause, excluding the pauses at turn-taking;
- relative duration: the duration of the last syllable divided by the average duration of the 2, 3, 4 and 5 previous syllables;
- relative pitch: the difference between the pitch (in semitones) of the last syllable and the average pitch of the 2, 3, 4 and 5 previous syllables;
- intra-syllabic pitch movement (in semitones)

The script also includes the information on the part-of-speech tag attributed to the corresponding token, and the corresponding expert annotation (by indicating whether the syllable marks the boundary of a specified unit).

The coding for prosodic units that will be used in the rest of the article is as follows: for the Rhapsodie corpus, four levels of annotation PER for periods, PCK for packages, GRP for groups and FT for feet; for the LOCAS-F corpus: B2 are boundaries of intermediate strength, B3 are strong boundaries, and HES indicate hesitations inhibiting the perception of a boundary. Syllables not marking a prosodic boundary are indicated by the symbol 0 (zero).

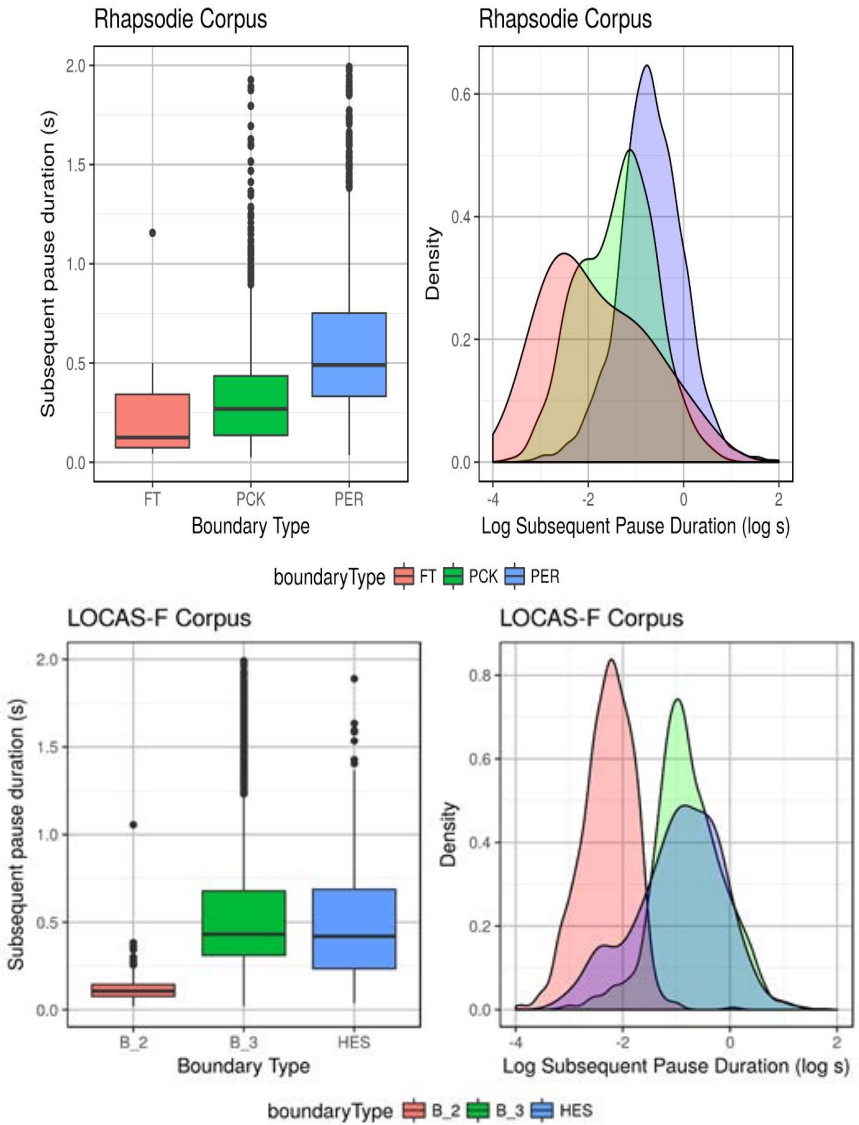
4 Results and discussion

In the following section, we will present the results of the statistical analysis of the measures extracted as described in the previous section, for each corpus.

4.1 Subsequent silent pause

The presence or absence of a silent pause immediately after a prosodic boundary appears to be the most important cue in distinguishing between boundaries of different strength (cf. also section 4.5 on the relative importance of the correlates). Figure 1 presents the distribution of the length of the subsequent silent pause for each type of prosodic boundary in each corpus. The original pause duration values have been used in the boxplots on the left; while the density distribution plots are based on the logarithmic transformation of pause duration. Since the typical distribution of pause durations is positively skewed, this transformation aims at approximating a normal distribution in log-time (see HELDNER; EDLUND, 2010 for a discussion of this method).

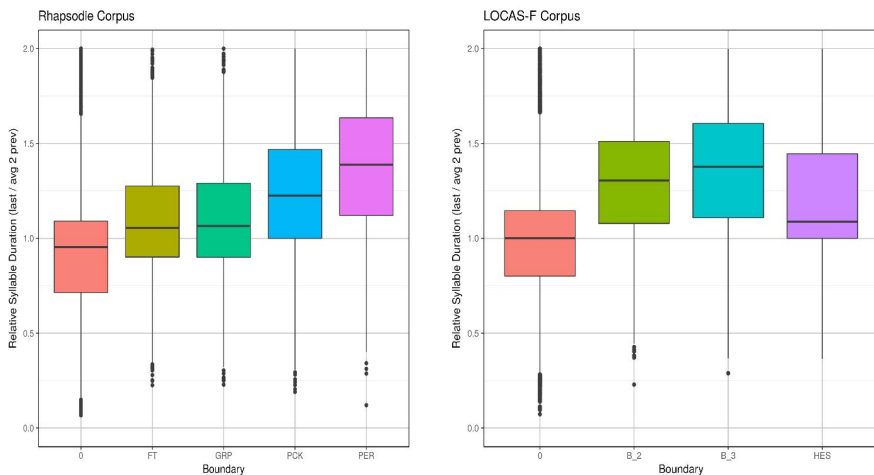
FIGURE 1 – Duration of the subsequent silent pause for each boundary type (feet, packages and periods in Rhapsodie; and B2, B3 and hesitations in LOCAS-F). On the left, the distribution is shown in seconds; on the right the duration has been log-transformed



4.2 Syllable lengthening

Syllables immediately preceding a prosodic boundary are often lengthened. We define the relative syllable duration as the ratio of the syllable duration at the unit end divided by the average duration of the previous two syllables. This ratio is a dimensionless quantity; a ratio of 1 indicates no lengthening, a ratio greater than 1 indicates lengthening and a ratio less than 1 indicates a local acceleration. Figure 2 shows the distribution of the relative syllable duration of the syllable immediately preceding each boundary type in each corpus. We observe that stronger prosodic boundaries are correlated with stronger syllable lengthening. In the Rhapsodie corpus, we observe that the last syllables of feet and groups are only slightly lengthened (it should be noted that syllable lengthening is also an acoustic correlate of syllabic prosodic prominence in French) and that the last syllables of packages and periods are lengthened. The pre-boundary syllable lengthening of packages in Rhapsodie is similar to the pre-boundary syllable lengthening of B2 boundaries in LOCAS-F, while the boundaries of periods in Rhapsodie correspond to the boundaries of B3 strength in LOCAS-F.

FIGURE 2 – Relative syllable duration (duration of the last syllable of a unit divided by the average duration of the previous 2 syllables) for each boundary type and corpus



4.3 Relative pitch and intra-syllabic pitch movement

In this section we will examine the intonation contours associated with prosodic boundaries. Figure 3 shows the distribution of the measure of relative pitch, defined as the difference between the mean pitch of the last syllable of a unit, and the average of the mean pitch of the preceding two syllables, in semitones relative to 1 Hz. These distributions are shown separately for prosodic boundaries with a rising intonation (relative pitch > 0) and a falling intonation (relative pitch < 0).

FIGURE 3 – Relative pitch (mean pitch of the last syllable of a unit minus the average of the mean pitch of the previous two syllables) for each boundary type and corpus. All pitch values are calculated on Prosogram-stylised syllables and are in semitones relative to 1 Hz

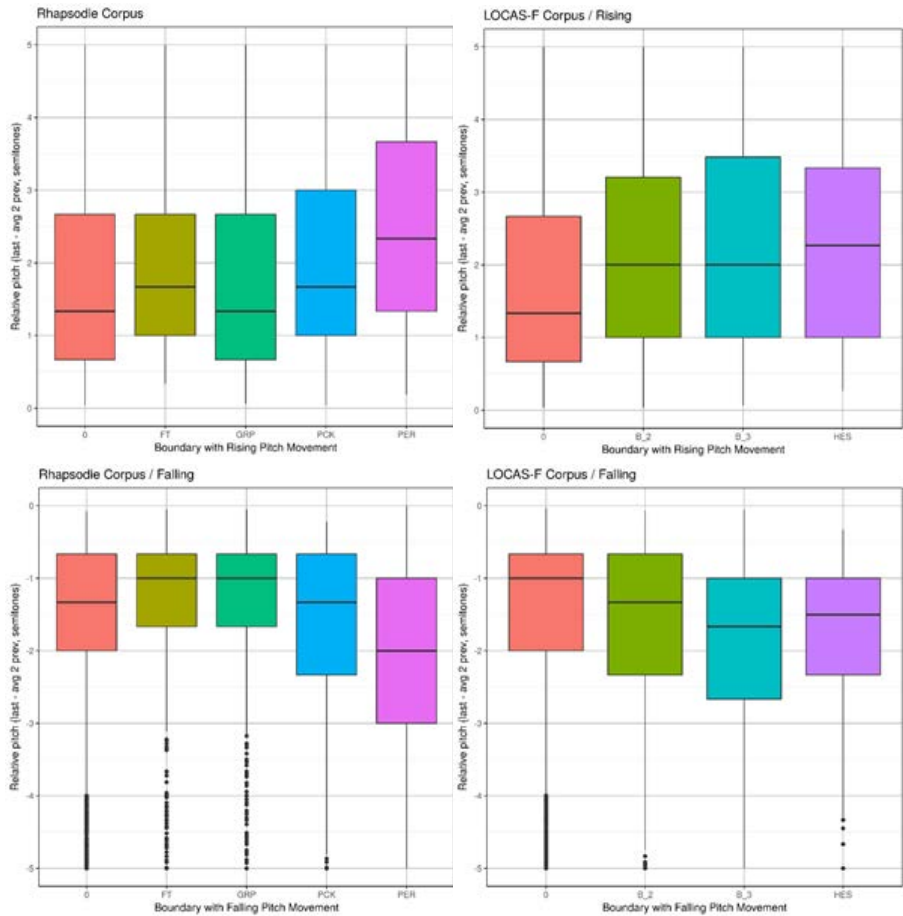
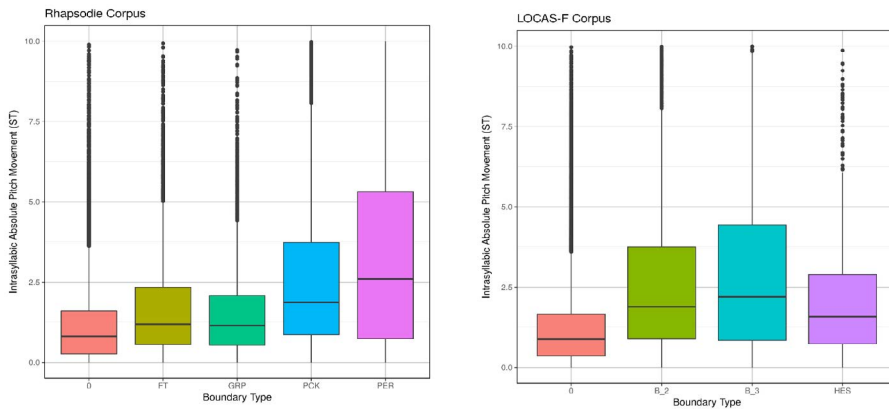


Figure 4 shows the distribution of the intra-syllabic pitch trajectory measure, i.e. the sum of absolute pitch intervals within syllabic nuclei divided by duration (in ST/s). A higher value indicates a syllable that will be perceived as more prominent, standing out of its neighbouring syllables. We observe that, in the Rhapsodie corpus, the last syllables of packages and periods have a significantly higher intra-syllabic pitch trajectory (with period-final syllables having a greater value than package-final syllables), while in the LOCAS-F corpus, the syllables associated with boundaries of both strengths (B2 and B3) have a higher trajectory than non-boundary syllables.

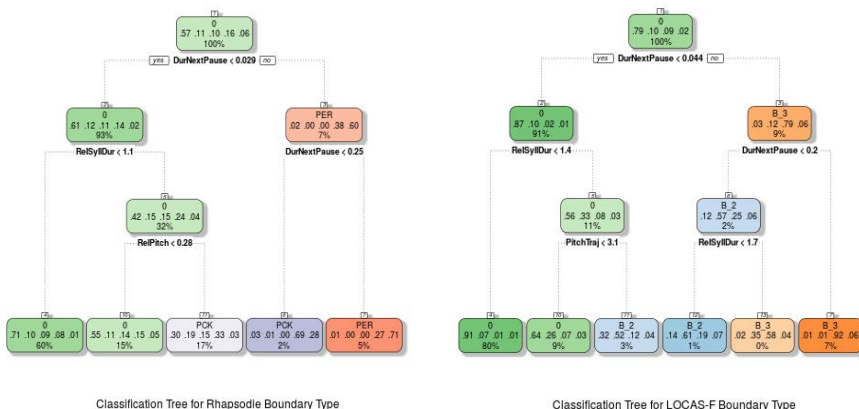
FIGURE 4 – Absolute intra-syllabic pitch movement (i.e. the sum of rising and falling intra-syllabic pitch movements, in semitones relative to 1 Hz)



4.4 Classification trees and relative importance of acoustic correlates

In order to evaluate the relative importance of each acoustic correlate in determining whether a syllable will be perceived as marking a prosodic boundary of a specific type, we calculated classification trees, using the `rpart` package, in the R statistical software system. The predictors for the classification algorithm were the acoustic correlates examined in the previous sections: the duration of the subsequent silent pause (if any), the relative duration of syllable compared to the previous two syllables, the relative mean pitch compared to the previous two syllables and the pitch trajectory. The resulting classification trees are shown in Figure 5.

FIGURE 5 – Classification Trees for each corpus and boundary type



Classification Tree for Rhapsodie Boundary Type

Classification Tree for LOCAS-F Boundary Type

We observe that the most important acoustic correlate for the perception of a prosodic boundary, in both corpora, is the subsequent silent pause duration. The next predictor, among the acoustic correlates, is the relative syllable duration, that effectively captures final lengthening of boundary syllables. Silent pause length and syllable lengthening distinguish between the presence and absence of a prosodic boundary and between boundary strengths (PCK and PER in Rhapsodie; B2 and B3 in LOCAS-F). Predictors related to pitch were found to be less important in both linear regression models: relative pitch distinguishes between no boundary and PCK boundary in the Rhapsodie corpus; and pitch trajectory distinguishes between no boundary and B2 boundary in LOCAS-F.

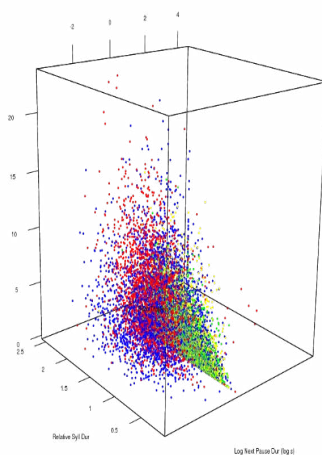
These corpus-based results on the acoustic correlates of prosodic boundaries are compatible with and confirmed by the series of experimental studies presented in Christodoulides *et al.* (2018). In this series of experiments, naïve listeners and expert annotators were asked to indicate the presence of a prosodic boundary in real-time, by tapping on a computer keyboard. The analysis of their responses, with a similar methodology (linear regression trees) shows that the most important correlate was the duration of the subsequent silent pause, followed by the co-occurrence of a major syntactic boundary, followed by final syllable lengthening and finally pitch movement. The relative importance of the

predictors is the same for the corpus-based analysis and the experiments, given that in the corpus-based analysis we are only considering signal-based (acoustic) correlates.

4.5 Clustering of different boundary types

Finally, Figure 6 presents a three-dimensional scatter plot, where each point corresponds to a syllable marking a prosodic boundary, of each of the four strengths defined in the Rhapsodie corpus. The points are colour-coded as follows: red represents period boundaries, blue represents package boundaries, green represents group boundaries, and yellow represents feet boundaries. The x axis is the log-transformed duration of the subsequent silent pause, the y axis is the relative syllable duration (as defined in section 4.2) and the z axis is the pitch trajectory (as defined in section 4.3). We observe that feet and group boundaries cluster together and that package and period boundaries cluster together, with period boundaries often being separated by way of the silent pause duration. This concurs with the results of the classification trees for the corpus.

FIGURE 6 - Scatter plot of syllable acoustic correlates for different types of prosodic boundaries in the Rhapsodie corpus. The boundaries are colour-coded as follows: periods – red; packages – blue; groups – green; feet – yellow.



5 Conclusion

In this article we have analysed two spoken French corpora, both containing samples from multiple speakers and speaking styles, and both having been annotated by experts for prosodic units and prosodic boundaries.

We have shown that the main acoustic correlates of prosodic boundary strength are the presence of a subsequent silent pause and pre-boundary lengthening, in this order of importance. Pitch movements (relative pitch and intra-syllabic pitch movement) are indicative of prosodic boundary function, rather than strength; however, stronger prosodic boundaries (e.g. period boundaries in the Rhapsodie corpus) tend to correlate with larger pitch movements.

With respect to our initial research question, on the relationship between two annotation systems for prosodic boundaries in French, which were developed independently from one another, we note that the two stronger boundary types in the Rhapsodie corpus are very similar to the intermediate and strong boundaries in the LOCAS-F corpus. Apart from its theoretical interest, this finding will facilitate the development of automatic annotation tools, by training machine learning models on the Rhapsodie corpus.

References

- ASTÉSANO, C., BERTRAND, R., ESPESSER, R.; NGUYEN, N. Perception des frontières et des proéminences en français. In *Actes des Journées d'études sur la parole et conférence annuelle du Traitement Automatique des Langues Naturelles*, Vol. 1: JEP, 353–360, 2012.
- AURAN, C., BOUZON, C.; HIRST, D. The AixMARSEC project: an evolutive database of spoken English. In *Proceedings of Speech Prosody 2004*, March 23–26, Nara, Japan, 2004, p. 561–564.
- AVANZI, M., SIMON, A. C., GOLDMAN, J.-P. & AUCHLIN, A. C–PROM: An annotated corpus for French prominence study. In *Proceedings of Speech Prosody 2010*, Prosodic Prominence Workshop, May 11–14, Chicago, USA, 2010.
- BOERSMA, P.; WEENINK, D. *Praat*: doing phonetics by computer [computer programme, version 6.0.36], Available at <http://www.praat.org>, 2017.

BRANCA-ROSOFF, S., FLEURY, S., LEFEUVRE, F.; PIRES, M. *Discours sur la ville*. Présentation du Corpus de Français Parlé Parisien des années 2000 (CFPP2000), 2012.

CHRISTODOULIDES, G. Praaline: Integrating Tools for Speech Corpus Research. In *Proceedings of the 9th International Language Resources and Evaluation Conference (LREC)*, Reykjavik, Iceland, 31–34, 2014. Available at: www.praaline.org.

CHRISTODOULIDES, G., AVANZI, M.; GOLDMAN, J.-P. DisMo: A Morphosyntactic, Disfluency and Multi-Word Unit Annotator: An Evaluation on a Corpus of French Spontaneous and Read Speech. In *Proceedings of the 9th International Language Resources and Evaluation Conference*, Reykjavik, Iceland, 3902–3907, 2014.

CHRISTODOULIDES, G.; SIMON, A. C. Exploring Acoustic and Syntactic Cues to Prosodic Boundaries in French: A Multi-Genre Corpus Study. In *Proceedings of ICPHS 2015*, Glasgow, Scotland, 2015.

CHRISTODOULIDES, G., SIMON A.C., DIDIRKOVÁ I. Perception of Prosodic Boundaries by Naïve and Expert Listeners in French. Modelling and Automatic Annotation. In: *Proceedings of the 9th Speech Prosody Conference*, Poznań, Poland, 13-16 June, 2018.

CLIFTON, C., CARLSON, K.; FRAZIER, L. Informative prosodic boundaries, *Language and Speech*, 45(2), 87–114, 2002. DOI: <https://doi.org/10.1177/00238309020450020101>

CRESTI, E.; MONEGLIA, M. *C-ORAL-ROM: Integrated reference corpora for spoken Romance languages*. Studies in corpus linguistics, 1388-0373. Amsterdam: John Benjamins, 2005. DOI: <https://doi.org/10.1075/scl.15>

CUTLER, A. Prosody in the comprehension of spoken language – A literature review. *Language and Speech*, 40(2), 141–201, 1997. DOI : <https://doi.org/10.1177/002383099704000203>

DI CRISTO, A. Vers une modélisation de l'accentuation du français: Première partie. *Journal of French Language Studies*, 9(2), 143, 1999. DOI: <https://doi.org/10.1017/S0959269500004671>

DI CRISTO, A. Une approche intégrative des relations de l'accentuation au phrasé prosodique du français. *Journal of French Language Studies*, 21(01), 73–95, 2011. DOI: <https://doi.org/10.1017/S0959269510000505>.

DURAND, J., LAKS, B.; LYCHE, C. *Phonologie, variation et accents du français*. IC2 : Trait  Cognition et traitement de l'information. Paris: Hermes / Lavoisier, 2009.

F RY, C. *Intonation and prosodic structure*. Key topics in phonology. Cambridge: Cambridge University Press, 2017. DOI: <https://doi.org/10.1017/9781139022064>

FRAZIER, L., CARLSON, K.; CLIFTON, C. Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, 10(6), 244–249, 2006. DOI: <https://doi.org/10.1016/j.tics.2006.04.002>

HELDNER, M.; EDLUND, J. Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4), 555–568, 2010. DOI: <https://doi.org/10.1016/j.wocn.2010.08.002>

JUN, S.-A.; FOUGERON, C. A phonological model of French intonation. In A. Botinis (Ed.) *Intonation: Analysis, Modelling and Technology*. Amsterdam: Kluwer Academic Publishers, p. 185–208, 2000. DOI: https://doi.org/10.1007/978-94-011-4317-2_10

LACHERET, A., KAHANE, S., BELI O, J., DISTER, A., GERDES, K., GOLDMAN, J.-P., OBIN, N., PIETRANDREA, P.; TCHOBANOV, A. Rhapsodie: A Prosodic – Syntactic Treebank for Spoken French. In *Proceedings of the 9th International Language Resources and Evaluation Conference*, Reykjavik, Iceland, 2014.

MARTIN, L. J., DEGAND, L., & SIMON, A. C. Forme et fonction de la p riph rie gauche dans un corpus oral multigenres annot . *Corpus*, 13(13), 243–265, 2014.

MERTENS, P. Intonational grouping, boundaries, and syntactic structure in French. In D. House & P. Touati (Eds.), *Proceedings of the ESCA Workshop on Prosody*, Vol. 41, 156–159, Lund (S) Working Papers. Lund University, 1993.

MERTENS, P. The Prosogram: Semi-Automatic Transcription of Prosody Based on a Tonal Perception Model. In B. Bel & I. Marlien (Eds.), *Proceedings of Speech Prosody 2004*, p. 549–552, 2004.

MERTENS, P. Polytonia: a system for the automatic transcription of tonal aspects in speech corpora. *Journal of Speech Sciences*, 4(2), 17–57, 2014.

MITTMANN, M. M.; BARBOSA, A. An automatic speech segmentation tool based on multiple acoustic parameters. *CHIMERA. Romance Corpora and Linguistic Studies, Madri*, v. 32, p. 133-147, 2016.

MO, Y., Duration and intensity as perceptual cues for naive listeners prominence and boundary perception. In *Proceedings of the 4th Speech Prosody Conference*, May 6–9, Campinas, Brazil, p. 739–742, 2008.

MO, Y., COLE, J. & LEE, E. Naive listeners prominence and boundary perception. In *Proceedings of the 4th Speech Prosody Conference*, May 6–9, Campinas, Brazil, p. 735–738, 2008.

OSTENDORF, M., PRICE, P. J.; SHATTUCK-HUFNAGEL, S. *The Boston University Radio News Corpus*, Boston University Technical Report No. ECS-95-001, March 1995.

PORTES, C. Approche instrumentale et cognitive de la prosodie du discours en Français. *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA)*, No. 21, p. 101–119, 2002.

ROSSI, M. *L'intonation: Le système du français : description et modélisation*. Collection L'essentiel français. Gap: Ophrys, 1999.

SCHUURMAN, I., SCHOUPPE, M., VAN DER WOUDE, T.; HOEKSTRA, H. CGN, An annotated corpus of Spoken Dutch. In A. Abbeilé, S. Hansen-Schirra, and H. Uszkoreit (Eds), *Proceedings of 4th International Workshop on Language Resources and Evaluation*, p. 340–347, 2003.

SMITH, C. Perception of prominence and boundaries by naïve French listeners. In *Proceedings of the 17th International Congress of Phonetic Sciences*, August 17–21, Hong Kong, China, p.1874–1877, 2011.

SWERTS, M. Prosodic features at discourse boundaries of different strength, *Journal of the Acoustical Society of America*, No. 101, p. 514–521, 1997. DOI: <https://doi.org/10.1121/1.418114>

WAGNER, M.; WATSON, D. G. Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, 25(7-9), p. 905–945, 2010, DOI: <https://doi.org/10.1080/01690961003589492>

WAGNER, P.; ORIGLIA, A.; AVEZANI, C.; CHRISTODOULIDES, G.; CUTUGNO, F.; D'IMPERIO, M.; ESCUDERO MANCEBO, D.; GILI FIVELA, B.; LACHERET, A.; LUDUSAN, B.; MONIZ, H.; NÍ CHASAIDE, A.; NIEBHUR, O.; ROUSIER-VERCRUYSSSEN, L.; SIMON, A.C.; SIMKO, J.; TESSER, F.; VAINIO, M. Different parts of the same elephant: A roadmap to disentangle and connect different perspectives on prosodic prominence. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*, August 10-14, Glasgow, UK, IPA, 2015. Available at: <http://hdl.handle.net/2078.1/161827>.

WATSON, D.; GIBSON, E. Intonational phrasing and constituency in language production and comprehension. *Studia Linguistica*, 59(2-3), p. 279–300, 2005. DOI: <https://doi.org/10.1111/j.1467-9582.2005.00130.x>



Automatic Speech Segmentation in French

Segmentação automática da fala em francês

Philippe Martin

LLF, UFRL, Université Paris-Diderot, Paris / France

philippe.martin@linguist.univ-paris-diderot.fr

Abstract: Whether we read aloud or silently, we segment speech not in words, but in accent phrases, i.e. sequences containing only one stressed syllable (excluding emphatic stress). In lexically stressed languages such as Italian or English, the location of stress in a noun, an adverb, a verb or an adjective (content words) is defined in the lexicon, and accent phrases include one single content word together with its associated grammatical words. In French, a language deprived from lexical stress, accent phrases are defined by the time it takes to read or pronounce them. Therefore, actual phrasing, i.e. the segmentation into accent phrases, depends strongly on the speech rate chosen by the speaker or the reader, whether in oral or silent reading mode. With a slow speech rate, all content words form accent phrases whose final syllables are stressed, whereas a fast speech rate could merge up to 10 or 11 syllables together in a single accent phrase with more than one content word. Based on this observation, and on other properties of stressed syllables, a computer algorithm for automatic phrasing, operating in a top-down fashion, is presented and applied to two examples of read and spontaneous speech.

Keywords: accent phrase; French; phrasing; stress location; boundary detection.

Resumo: Quando lemos em voz alta ou silenciosamente, segmentamos a fala em palavras, mas em grupos acentuais, i.e., sequências contendo uma única sílaba acentuada (excluindo-se acento enfático). Em línguas lexicalmente acentuadas como o italiano ou o inglês, a localização do acento em um substantivo, um advérbio, um verbo ou em um adjetivo (palavras lexicais) é definida no léxico, e sintagmas acentuais incluem uma única palavra lexical, acompanhada das palavras gramaticais a ela associadas. Em francês, uma língua que não possui acento lexical, sintagmas acentuais são definidos pelo tempo que se leva para lê-los ou pronunciá-los. Assim, os constituintes concretos,

i.e., a segmentação em grupos acentuais, depende fortemente da velocidade de fala escolhida pelo falante ou leitor, tanto na fala como na leitura silenciosa. Com uma velocidade de fala baixa, todas as palavras lexicais formam grupos acentuais cujas sílabas finais são acentuadas, enquanto o ritmo de fala rápido poderia juntar de 10 a 11 sílabas em um mesmo grupo acentual contendo mais de uma palavra lexical. Com base nessa observação e em outras propriedades das sílabas acentuadas, um algoritmo computacional para segmentação automática, atuando de maneira top-down é apresentado e aplicado a dois exemplos de leitura e fala espontânea.

Palavras-chave: grupo acentual; francês; segmentação; posição do acento; detecção de fronteira.

Submitted on January 12th, 2018

Accepted on February 27th, 2018

1 Introduction

When we read a text, either aloud or silently, we could proceed word by word, or even syllable by syllable, but if we master the language and identify all the words, we usually proceed by group of words. It is easy to observe in an orthographic transcription where all words would be ended by a final dot that we don't read word by word, as it would be the case in: *In. the. Orthographic. Representation. Of. Speech. Of. Most. Written. Languages. Segmentation. Is. Defined. By. Spaces. Between. Words.* Instead, we normally read a sentence by grouping words in units containing either a noun, an adverb, a verb or an adjective (i.e. a content word), together with the grammatical words (pronoun, conjunction...) associated to them to form an *accent phrase*. The preceding example, segmented in accent phrases, indicated in squared brackets, would be: [*in the orthographic*] [*representation*] [*of speech*] [*of most*] [*written*] [*languages*] [*segmentation*] [*is defined*] [*by spaces*] [*between*] [*words*]. Each of these groups carry a single stressed syllable placed on some syllable of the content words as defined in the lexicon of English: [*in the orthographic*] [*representation*] [*of speech*] [*of most*] [*written*] [*languages*] [*segmentation*] [*is defined*] [*by spaces*] [*between*] [*words*]. Such groups of words are called in prosodic phonology *accent phrases*, and define the minimal prosodic units, which organized into a hierarchy, constitute the *prosodic structure* of the sentence (MARTIN, 1975, SELKIRK, 1978).

For all fluent speakers of English, the position of stressed syllables in accent phrases is predictable, and results from the acquisition of the lexicon of the language. Other stressed syllables can also occur in speaker's production, but contrary to lexically defined stress, they are not predictable as they result from a specific choice of the speaker to indicate an emphasis, as in *segmentation in most written languages* with a stress on the first syllable of *segmentation*. This kind of emphatic stress may occur on a different syllable than the lexically stressed or on the same syllable. In this latter case, the speaker will use a different acoustic realization, as emphatic stress has to be perceived by listeners as different and unpredictable compared to the predictable lexical stress.

The predictability of lexical stress suggests that the perception of stressed syllables may not be directly derived from the processing of specific acoustic features of speech, such as vowel duration, fundamental frequency change or intensity modulation, the prosodic parameters often mentioned in the literature as parameters of stress. Instead, the perception of stressed syllables could be considered as the result of an identification mechanism comparing the actual acoustic features of syllables with a predicted position derived from the knowledge of the language. As in silent reading as well as reading aloud, segmentation into accent phrases is inevitable, the same process takes place when we listen to somebody speaking, eventually restoring stress in a position where we would have placed the stressed syllable ourselves.

One can mention on this topic the experiment on the perception of accented syllables of Berber and Hebrew by subjects who had no notion of these languages at all (METTOUCHI *et al.*, 2007). The acoustic features are present in the speech signal, but in this experiment the listeners didn't identify any stress locations (except by chance...), positioning stress on syllables belonging to sequences they thought they had identified through the perception grid of their mother tongue (or another they knew). Indeed, no appropriate lexicon allowing the listeners to position an expected stressed syllable and interpret the acoustic data was available, which is not the case for speakers of Berber or Hebrew. Similar observations can be found in Astésano and Bertrand (2016) and Michelas *et al.* (2016).

2 The case of French

French is a language where the position of lexical stress evolved gradually to the last syllable of content words (actually to the last syllable of any word pronounced in isolation) by progressively dropping all post-stressed syllables (VÄÄNÄNEN, 1995). The function of lexical stress as marker of morphological boundary as in lexically stressed languages was gradually lost as redundant. Since its main phonological function was lost, it became then possible for speakers to skip some of the predicted stress locations when speaking or reading. This can be seen in *la ville de Versailles* (“the city of Versailles”); which can be read with one or two stressed syllables, placed on the last syllable of content words *ville* and *Versailles*: *la ville de Versailles* or *la ville de Versailles*. Likewise, an example such as *la petite armoire violette* (“the little purple cupboard”), can receive one, two or even three stressed syllables: *la petite armoire violette*, *la petite armoire violette*, *la petite armoire violette* or *la petite armoire violette*. For a French speaker, it is easy to realize that the difference in phrasing of these examples is linked to the speech rate, possibly leading to a different processing of the sentence content. In order to pronounce (or even to read silently) *la petite armoire violette* with only one final stressed syllable on *violette*, one has to use a (very) fast speech rate, whereas a slower pace would lead to the pronunciation of three stressed syllables as in *la petite armoire violette*. Surprisingly, this dependency of phrasing to the speech rate seems to escape some researchers who are native speakers of French, as it appears in a recent issue of the review *Langue Française* (2016, n. 191), gathering papers devoted to *accentuation et phrasé*. The absence of the time parameter implied in phrasing even lead to the often-mentioned belief that French listeners are ‘deaf’ to stress...

We could perhaps then conclude that there is no limit to the number of syllables and thus of words that can be pronounced in French with only one final stressed syllables, and that can be inserted in a single accent phrase. The pronunciation of long words will help discover where the limit stands. Long words such as the well-known *anticonstitutionnellement* (“against the constitution”), (8 syllables) or *intergouvernementalisation* (“inter governmentalization”) (10 syllables) seem difficult if not impossible to pronounce or read even silently with only one final stressed syllable. Already in the 16th century, the

grammarians Louis Meigret (1550) proposed that the longest word that could be pronounced with only one final stress would have a maximum of 7 syllables. Much later, Martin (2014) showed that it was not the number of syllables that matters, but the time needed to pronounce them, even in silent reading. The data obtained from fast speech rate speakers suggest that the maximum interval between consecutive stressed syllables (in flowing speech) could not exceed some 1,250 ms, depending on the subjects. In *parler jeune* productions in French (the young people speaking style), sequences of up to 10 or 11 syllables with only one final stress have been observed (LEKHA; LE GAC, 2004). This value is close to the theoretical limit, derived from the minimal average duration of syllables that could be perceived in a sequence, about 100 ms (GHITZA; GREENBERG, 2009). These observations would put the maximal duration of accent phrases in French to about 1,250 ms to 1,400 ms or so, with the fastest speech rate reaching about 11 or 12 syllables per second.

If 1,250 ms (the approximative value retained in this paper) is the maximal duration between consecutive stressed syllables in connected speech, there is also a minimal duration that exists between two consecutive stressed syllables. This value will define a minimal duration for accent phrases that would contain only one syllable. Its value is experimentally easy to evaluate, by selecting natural or synthetic occurrences of consecutive stressed syllables, as for example *par le **fait que*** (“by the fact that”) or *le travail de **nuît nuît*** (“night work harms”) i.e. cases of stress clash with no move or deletion of the first stress. It is often mentioned in the literature that these cases require a kind of acoustic gap between consecutive stressed syllables (e.g. DI CRISTO, 2016), usually but not always implemented by the presence of consonants after the first of before the second stressed syllable (which is the case for the two examples above). However, it is easy to experimentally reduce the gap with a sound editor until the first implied syllable ceased to be perceived as stressed although nothing of its acoustical structure has been modified (i.e. by removing the silent part only). This limit is about 250 ms (depending on the way distances are measured between syllables, from their center or from the two third of their duration), which gives the minimal duration of an accent phrase, since below this value, the word owning the first syllable will become part of the newly formed accent phrase. For example, the perceived desaccentuation of *fait* in [*par le **fait***] [***que***] (“by the fact that”) will merge the accent phrase *par le **fait*** with

the second accent phrase *que* to form the new group [*par le fait que*]. The minimal duration between two consecutive stressed syllables is thus about 250 ms (MARTIN, 2014), which implies that a one-syllable accent phrase must include some voiceless or silent segment that precedes it, as the preceding vowel, if exists, is necessarily stressed and ends the preceding accent phrase.

3 Syllables followed by silence are stressed in French

It is equally easy to demonstrate experimentally that any syllable followed by at least 250 ms of silence is perceived as stressed in French. The fact that any final syllable is perceived as stressed is a consequence of the prepositioning of the stressed syllable by the listener, as final syllables are stressed in French, and that a silent gap following the end of an accent phrase is necessarily stressed.

Either by inserting some 250 ms or more acoustic silence on the speech wave, using a sound editor without modifying the acoustic characteristics of the syllable at all, or by simply slowing the speech rate so that the number of syllables reaches a level below some four syllables per second, the final syllables of any word category becomes perceived as stressed, whatever their actual duration or pitch movement. In lexically stressed languages, the perception of an accent phrase final syllable as stressed is preempted by the position of lexical stress (if not in final position). In Italian for example, the lexical stress of the penultimate syllable of *Marco* in *la sorella di Marco è partita* (“Marco’s sister left”) prevents a listener who knows the language to perceive the last syllable *co* as stressed, although it is followed by more than 250 ms of silence, whereas a speaker of French who does not know Italian will perceive the final syllable of *Marco* as stressed.

The important parameter in these cases pertain to the lack of speech data to be processed by the listener and the actual explanation is linked to the processing of syllables by the brain, and more precisely by the brain oscillations carrying information between neuronal zones (MARTIN, 2015). As mentioned above, it can be shown that the perception of syllables needs at least 100 ms processing time, even if their actual duration is below this value. If given more than some 250 ms, a normally unstressed syllable becomes perceived as stressed, without modification of its acoustic structure. Since two consecutive stressed

syllables must be separated by at least 250 ms, we can conclude that the perception of *stressed* syllables needs at least 250 ms processing time. This is a consequence of the processing of stressed syllables by delta brain waves (MARTIN, 2018).

In summary, a normally unstressed syllable can be perceived as stressed by timing characteristics pertaining to a silent gap following the syllable itself. Likewise, a normally stressed syllable can be perceived unstressed for a similar reason, the gap duration existing between two consecutive syllables.

4 Pronouns

The *pronoms toniques* in French (*moi, toi, lui, elle, nous, vous, eux, elles*) do not belong to the category of content words, but share their characteristics in term of accent phrase stress, in particular in examples with a tonic pronoun placed after the verb. The normal stress pattern of *redonne moi la main*, [*redonne*] [*moi la main*] (“give me your hand again”) leads to the unexpected accent phrase [*moi la main*], *redonne moi la main* being emphatic, the stress pattern [*redonne*] [*la moi*] [*plus loin*] (“give me it further”) is quite possible and leads to consider some tonic pronouns as stressable even if they are not followed by 250 ms of silence. There are cases where tonic pronouns are effectively stressed, although they do not belong to the content word category. In other configurations, as in *moi ma mère le salon c’est de la moquette*, the tonic pronoun *moi* is stressed if followed by 250 ms of silence, *moi # ma mère le salon c’est de la moquette* (“me my mother the living room is carpet”), but unstressed if there is no sufficient gap after *moi*, as predicted: *moi ma mère...* The same configuration can be observed in well-known examples such as *mon manège à moi c’est toi* (“my ride to you is me”), from a famous Edith Piaf song, or *Je est un autre* (“I am another”), Arthur Rimbaud.

Likewise, demonstrative pronouns are also stressable although they don’t belong to the content word category. In *...pour tous ceux et toutes celles...* (NS) “for all those...”, both demonstrative pronouns are stressable and stressed. The same observation applies to possessive pronouns such as *le mien, le tien, la leur, les leurs*.... “mine, yours, their, theirs”.

Finally, relative pronouns (*qui, que, quoi, dont, où, lequel...*) are also stressable, but become stressed mainly if followed by a 250 ms silence.

5 Eurhythmmy

The eurhythmicity observed for both read and spontaneous speech may also be taken into account in a top-down approach for prosodic segmentation, i.e. an approach not processing from acoustic data to phonological conclusions, but rather from the general properties of stressed syllables to eventually validate acoustic data in the speech signal. As a general observation (WIOLAND, 1985), spontaneous speech eurhythmmy proceeds by adjusting the average duration of accent phrases syllables to reach comparable duration of successive accent phrases. Read speech uses more often a strategy aiming to balance the number of syllables of successive accent phrases, at the possible expense of congruence with the syntactic structure. A classic example is given by a sentence such as *Marie adore les chocolats* (“Mary loves chocolates”) in which spontaneous speech subjects would have a tendency to realize a phrasing congruent with syntax [*Marie*] [*adore les chocolats*] and possibly aim for eurhythmmy by slowing the syllabic rate of [*Marie*] and going faster on [*adore les chocolats*]. On the contrary, readers of this sentence show a tendency to group the words to balance the number of syllables in consecutive accent phrases, at the expense of congruence with syntax [*Marie adore*] [*les chocolats*].

To implement eurhythmicity in a segmentation algorithm, an average duration of accent phrases can be estimated in a running window containing some 3 or 4 consecutive accent phrases. This value should be between about 250 ms (each syllable is followed by 250 ms silence, a production style where all syllables are pronounced detached) to about 1250 ms, characteristic of the *parler jeune*. Assuming the speaker or reader rhythm does not vary too much in a given amount of time, a more or less reliable duration value is obtained from two or three consecutive accent phrases duration values. Experimental data obtained from spontaneous speech show that the average accent phrase duration is about 500 to 700 ms (MARTIN, 2018).

To summarize the properties and observations on accent phrase stress in French:

1. Duration of accent phrases (in French): between 250 ms and 1,250 ms;
2. Accent phrases may contain 1 to 11 syllables ;
3. The minimal and maximal speech rates are between 4 and 10 syllables per second (in continuous speech);
4. Any syllable followed by more than 250 ms silence is perceived as stressed;
5. Eurhythmicity aims to balance the duration of successive accent phrases.

6 Virtual and actual stress, stressable and stressed syllables

Phrasing determines an essential step in the comprehension of speech. The segmentation into accent phrases constitutes the first phase to rebuild the prosodic structure intended by a speaker, which is essential and unavoidable to access the syntactic structure when we read. The resulting prosodic structure will not necessarily match the prosodic structure intended by the writer of the text we read, which leads to consider reading resulting from our own segmentation of the text, as the phrasing depends on the reading speed selected, and this is true in both reading aloud or silently. The only limits to these variations are given by the minimal and maximal duration of accent phrases.

The simple fact that we can restore stress locations when we read aloud or silently tells us that we may not really need any acoustical input to perceive stressed syllables (again non-emphatic). Not only reading aloud or silently of the same text could lead to different phrasings, but while listening to speech, we cannot prevent to have expectations towards the location of stressed syllables different from the one actually realized by the speaker. In other words, we can “hear” stressed syllables that actually may not be present acoustically. This apparent illusion is a direct consequence of many perception processes in speech (ARNAL; GIRAUD, 2017) involving not a direct processing of some physical input, but rather the validation of an expected input by comparison between what’s expected and what is actually physically realized. In the case of accent phrase defined by a final stress, we can predict from our lexicon the location of a stressed syllable in a group of words, which will depend on the speech rate selected in this operation. Considering again the former example *la petite armoire violette*, the speaker could have stressed *armoire* and *violette*, *la petite armoire violette*, but we may

have expected a slower speech rate and mentally also stressed *petit: la petite armoire violette*. Consequently, we could then hear three stressed syllables although the speaker had realized only two. The only way to avoid this perception of this *virtual* stress (for syllables that would be stressed), opposed to the *actual* stress (for effectively stressed syllables) present in the acoustic wave, would be to constantly adapt our speech rate to the one used by the speaker, or the one assumed to be used by the scripter.

This adaptation is not always easy or even possible. The examples provided by the *parler jeune* with a very fast speech rate exceeding 7 or 8 syllables per second are hard to match for most listeners, to the point that some will have trouble to understand such speech tempo. Therefore, some listeners will have a tendency to hear stressed syllables where they do not exist acoustically. In the example illustrated in Fig. 1 displaying a speech wave and the corresponding fundamental frequency curve, the actual accent phrases acoustically realized by the speaker are

[*C'est toi qui a pris la responsabilité de casser*] pronounced with 13 syllables

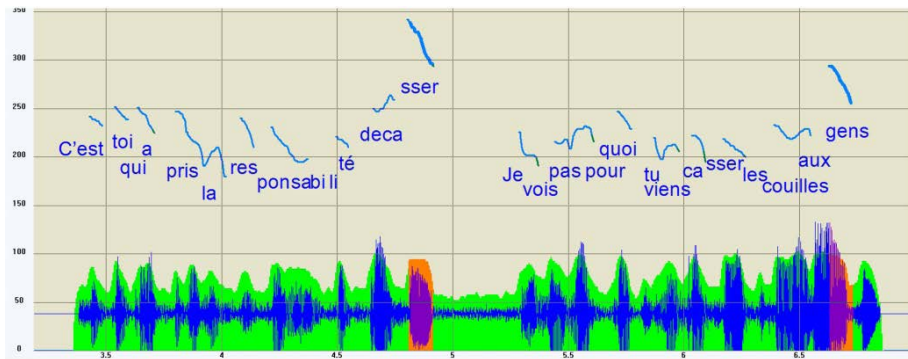
[*Je vois pas pourquoi tu viens casser les couilles aux gens*] 12 syllables

(“you took the responsibility to break up. I do not see why you come to break people’s balls”). (corpus *l'Esquive*)

The first accent phrase contains 13 syllables, which is unusual for the average speaker of French. Therefore, any listener not practicing the *parler jeune* will have a strong tendency to restore mentally a stressed syllable on *toi* leading to a different phrasing [*C'est toi*] 2 syllables [*qui a pris la responsabilité de casser*] 11 syllables (elision of [i] in *qui a*), or even also on *pris*, resulting in a four accent phrases phrasing [*C'est toi*] 2 syllables [*qui a pris*] 2 syllables pronounced [kapri] [*la responsabilité*] 7 syllables, [*de casser*] 3 syllables. Still, as shown on Fig. 1, the only obvious acoustical marker of stress is on the final syllable of *casser*.

Likewise, the second accent phrase with 12 syllables, could be mentally segmented into 2 or 3 accent phrases, depending on the speech rate adopted mentally or in oral production: [*Je vois pas pourquoi*] [*tu viens casser les couilles aux gens*] (the [ə] of the pronoun *je* is deleted here), or [*Je vois pas pourquoi*] [*tu viens casser*] [*les couilles aux gens*].

FIGURE 1 – An example of fast speech rate with 13 and 12 syllables in the accent phrases: *C’est toi qui a pris la responsabilité de casser. Je vois pas pourquoi tu viens casser les couilles aux gens* («you took the responsibility to break up. I do not see why you come to break people’s balls”).



The possible discrepancy between perceived and realized stressed syllables in French leads to differentiating virtual from actual stressed syllables. Virtual syllables correspond to what Paul Garde (1968, 2013) called *stressable*, whereas the syllable effectively marked by acoustic parameters are then *stressed*. The number of stressable syllables is necessarily equal or superior to the number of actually stressed ones, whose number depends on the speech rate.

7 Stress annotation: *mission impossible* ?

The problem for an annotator of stressed syllables (outside emphasis) in French is to adapt to the speech rate of the recording when accented syllables are annotated. The perception of stress will be influenced by the annotator’s own prediction process, thereby tending to detect stressed syllables where they would have been placed by reading or speaking not at the speaker’s speech rate but at the annotator’s own pace.

Most often, implemented automatic detection of stressed syllables in French operates in a bottom-up fashion from the speech recording, looking for significant variations between consecutive syllables in duration, fundamental frequency and intensity (for recent examples, see GOLDMAN *et al.*, 2013; MERTENS; SIMON, 2013). Vowel quality does not appear as a significant parameter for stress detection in French. Still, top-down approaches do exist, essentially applied to English (ARNOLD; WAGNER, 2008), operating from the word category to detect syllabic stress.

In a paper published in 2013, M. Avanzi, faced with the uncertainty of annotating stressed syllables in French, describes in detail a complex procedure involving two experts, possibly helped by a third in case of disagreement between the first two. Even with this protocol, agreement between annotators varies between 60 % and 80 %. In Martin (2006), some explanations were already proposed for the lack of convergence observed in the perception of stressed syllables in French by experts. These explanations pertain to the expectations of stress placement by various annotators, experts or non-experts.

In another paper on the same topic, Christodoulides and Avanzi (2014) implemented an automatic detector of prominence (i.e. not just accent phrase stressed syllables) by machine learning methods applied to a large corpus (11 hours) which included two different speech styles. They use a comprehensive set of acoustic parameters that they hoped would be appropriate to differentiate prominent syllables from others (syllabic duration, minimal, maximal average fundamental frequency, pitch movement, peak intensity, spectral balance, part of speech tag, presence and duration of subsequent pause, syllabic structure, position of the syllable in the word). Their best results, evaluated against manual placement by experts in syllabic prominence (therefore subject to the limitations evoked above), reaches a 90% correct identification level.

Considering these difficulties, it appears that stress detection should proceed not from the speech wave analysis, but rather from the knowledge a system could have access to beforehand, the location of potential positions as final syllables of content words among others, i.e. proceed in a top-down fashion.

Indeed, as we have seen above, the perception of stressed syllables by listeners proceeds not by direct evaluation of actual acoustic parameters in the speech wave, but rather by comparison of listener expected stress locations with perceived acoustic parameters. In this process, the evaluation of an expected stress position and actual realization by the speaker precedes the actual validation process comparing expectation and reality. This explains why even expert listeners may perceive as stressed syllables not carrying specific acoustic features differentiating from surrounding syllables, and how we restore stressed syllables in silent reading without any actual acoustic information.

To attain a reasonable chance of success, a computer implementation dealing with speech wave should then adopt a comparable

strategy, and not infer results starting from acoustical analysis of the speech wave but rather from expectation of stressed syllable locations. The availability of transcribed and segmented speech data, down to the syllabic and phone level, should be a prerequisite towards automatic stress detection, as the candidates for syllabic stress can be directly inferred from the aligned transcription.

8 A top-down algorithm

To apply the definition given in lexically stressed languages to French, we can assign a virtual stress to final syllables of all words belonging to the category of noun, verb, adverb, adjective and pronoun. To help select actual stressed syllables among the list of stressable ones, we can in a first step use the constraints described above, i.e. the minimal and maximal duration of accent phrases (respectively 250 ms and 1,250 ms), the minimal separation of 250 ms between two consecutive stressed syllables, and the presence of at least 250 ms of silence following a virtual stress. The application of these constraints would make some virtual candidate stressed syllables actually stressed (as a unique stressable syllable) in a time window of 1,250 ms for example, and eliminate some from the list of possible actual stressed syllables (the first stressable syllable cannot be actually stressed if closer to the next stressed syllable by less than 250 ms).

As stated above, the next step to select stressed syllables effectively without even starting accent phrase looking at the speech wave would be to look at the speech rate, i.e. the number of syllables per second actually observed on the transcription of the speech wave. Linked to an average number of syllables per accent phrase, we can then have an approximation of the phrasing realized in a given recording, validated by an assumed eurhythmicity.

To finally exploit the actual acoustic data, and innovate from the existing list of traditional parameters, i.e. changes / contrasts in syllabic duration, fundamental frequency and frequency, we could refer to the function of syllabic stress to define accent phrases as minimal units of the sentence prosodic structure. According to the model of Martin (1975, 1987), the prosodic structure results from a dynamically built hierarchical organization of accent phrases. From the presence of an expected terminal conclusive contour, perceived as a marker of non-

continuation of the sentence, two other melodic contours, one rising, the other falling, indicate respectively a major and a minor continuity.

The interesting characteristics of the continuity contours (always located on the vowel of the stressed syllable) is that they indicate a dependency relation, minor continuation towards major continuation and major continuation towards the terminal conclusive contour, by a contrast of melodic slope, where a falling contour indicates a dependency toward a rising contour. Of course, this model implies that the falling and rising melodic slope are effectively perceived, i.e. that the speed of melodic change in time is above what is called the glissando threshold. The glissando threshold is evaluated as the difference from the beginning to the end in semitones referred to the duration of the contour (assuming a linear variation, cf. ROSSI, 1971).

According to this definition of accent phrases as minimal units of prosody whose hierarchy constitute the sentence prosodic structure, we can designate any stressable syllable whose change in fundamental frequency on its vowel exceeds the glissando threshold as effectively stressed. Although this step assumes the validity of the glissando threshold (which in fact implies an adjustment parameter), as well as the linearity of the fundamental frequency change of the syllable vowel used for the evaluation of the glissando value, we have enough tools to implement an innovative algorithm for automatic selection of stressed syllables from a list of stressable syllables.

9 Automatic detection of stressed syllables in French

From these various observations and considerations, the following rules for a computer implementation can be applied:

1. Any syllable followed by more than 250 ms silence is stressed;
2. Any final syllable of a noun, adjective, verb, adverb or pronoun is stressable (from accent phrase definition);
3. If two consecutive stressed syllables are separated by less than 250 ms, the first one is unstressed (accent phrase minimum duration from the minimum spacing between consecutive stressed syllables);
4. Any stressable syllable with change of fundamental frequency over the glissando threshold is stressed;

5. If two consecutive stressed syllables are separated by more than 1,250 ms in continuous speech, at least one stressable syllable in this interval is stressed (accent phrase maximum duration). Make stressed the one with the highest glissando value;
6. One stressable syllable must exist in any time window duration equal to the accent phrase average duration (eurhythmy).

The eurhythmic aspect is implemented by evaluating the first accent phrases realizations and the number of syllables they contain. This starting accent phrase duration will then be used to define a sliding time window, in which most prominent syllables in value of glissando will be retained as stressed. The size of this sliding window defines a speech rate assumed to be constant in the whole recording.

10 An example of read speech

A first read example: *il était une fois un pauvre escargot qui souffrait beaucoup à chaque fois qu'il partait en randonnée car il avait du mal à suivre le rythme de ses compagnons* ("Once upon a time, there was a poor snail who suffered a lot every time he went on a hike because he had trouble keeping pace with his companions").

In the steps detailed below, stressable syllables are underlined, and stressed syllables are underlined and bold.

Step 1: Any syllable followed by more than 250 ms silence is stressed:

*Il était une fois un pauvre escargot qui souffrait beaucoup à chaque fois qu'il partait en **randonnée***

Step 2: Any final syllable of a noun, adjective, verb, adverb or tonic pronoun is stressable:

*Il était une fois un pauvre escargot qui souffrait beaucoup à chaque fois qu'il partait en **randonnée***

Step 3: If two consecutive stressed syllables are separated by less than 250 ms, the first one is unstressed: the gap between *chaque* and *fois* is 180 ms, below the 250 ms limit:

*Il était une fois un pauvre escargot qui souffrait beaucoup à chaque |180 ms| fois qu'il partait en **randonnée***

Step 4: Any stressable syllable with F0 change over the glissando threshold is stressed {glissando value/glissando threshold with coefficient 0.16}.

The stressable syllables below the threshold are unstressed:

Il était {35/76} *une fois* {36/17} *un pauvre* {44/66} *escargot* {32/12}
qui souffrait {54/144} *beaucoup* {79/66} *à chaque fois* {46/106} *qu'il*
partait {32/51} *en randonnée*

Step 5: Two consecutive stressed syllables separated by more than 1,250 ms, as in the case of the last accent phrase:

[à chaque fois qu'il partait en randonnée] 1367 ms

We can select the highest glissando value, on *fois*:

[à chaque fois qu'il partait en randonnée] 1367 ms

or both stressable syllables on *fois* and *partait*:

[à chaque fois qu'il partait en randonnée] 1367 ms

Step 6: Apply eurhythmicity to retain the latter possibility:

726 ms 5 syl. 145 ms/syl. *Il était une fois*

687 ms 5 syl. 137 ms/syl. *un pauvre escargot*

765 ms 5 syl. 153 ms/syl. *qui souffrait beaucoup*

407 ms 3 syl. 135 ms/syl. *à chaque fois*

487 ms 3 syl. 162 ms/syl. *qu'il partait*

546 ms 4 syl. 136 ms/syl. *en randonnée*

The average accent phrase duration is about 709 ms.

11 An example of spontaneous speech

The second example belongs to the category of *parler jeune*: *Juste pour une carte d'identité t'as pas ta carte tu fais tes vingt-quatre heures tu ressorts t'as la haine encore plus ça augmente* (“Just for an identity card you do not have your card you make your twenty-four hours you come out you hate even more it increases”).

Step 1: The last syllable is followed by more than 250 ms of silence:

Juste pour une carte d'identité t'as pas ta carte tu fais tes vingt-quatre heures tu ressorts t'as la haine encore plus ça augmente

Step 2: Any final syllable of a noun, adjective, verb, adverb or tonic pronoun is stressable:

Juste pour une carte d'identité t'as pas ta carte tu fais tes vingt-quatre heures tu ressors t'as la haine encore plus ça augmente

Step 3: If two consecutive stressed syllables are separated by less than 250 ms, the first one is unstressed: the gap between *encore* and *plus* is 240 ms, below the 250 ms limit:

Juste pour une carte d'identité t'as pas ta carte tu fais tes vingt-quatre [230 ms] heures tu ressors t'as la haine encore [240 ms] plus ça augmente

Step 4: Any stressable syllable with F0 change over the glissando threshold is stressed. The stressable syllables below the threshold are unstressed:

Juste {64/36} *pour une* carte {44/38} *d'identité* {54/45} *t'as pas* ta carte {44/38} *tu fais* {54/142} *tes vingt-quatre* [230 ms] heures {49/37} *tu ressors* {38/32} *t'as la* haine {25/22} *encore* [240 ms] plus {38/23} *ça augmente*

Steps 5 and 6 do not apply:

227 ms 1 syl. 227 ms/syl. *Juste*

356 ms 3 syl. 118 ms/syl. *pour une carte*

537 ms 4 syl. 134 ms/syl. *d'identité*

569 ms 4 syl. 142 ms/syl. *t'as pas ta carte*

945 ms 6 syl. 157 ms/syl. *tu fais tes vingt-quatre heures*

486 ms 3 syl. 162 ms/syl. *tu ressors*

431 ms 3 syl. 143 ms/syl. *t'as la haine*

496 ms 3 syl. 165 ms/syl. *encore plus*

592 ms 3 syl. 197 ms/syl. *ça augmente*

The average accent phrase duration is 515 ms.

12 Conclusion

Contrary to lexically stressed languages such as English or Italian, in which accent phrases contain one stressed syllable usually carried by a

content word, French segmentation into accent phrases depends strongly on the speaking or reading rate used. In fact, the only limitation for the number of words contained in a single accent phrase in French is the time taken to pronounce them, which cannot exceed some 1,250 ms, even in silent reading or speaking to oneself.

In view of this property, and of the fact that the perception of stressed syllable results from a validation process comparing the predicted position with the actual acoustic parameters, a top-down automatic phrasing segmentation in French is briefly described. The algorithm incorporates the following observations: 1) Speakers and readers of French are capable to restore accent phrase stressed syllables even without any acoustic input; 2) The minimum duration of accent phrases is 250 ms, and the maximum about 1,250 ms; 3) The actual duration of accent phrases depends on the speech rate selected by the speaker or the reader; 4) The actual syllabic stress defining phrasing carries a melodic movement above the glissando threshold.

References

ARNAL, L. ; GIRAUD, A-L. Neurophysiologie de la perception de la parole et multisensorialité. In : PINTO, Serge ; SATO, Marc (Ed.). *Traité de neurolinguistique*. Louvain-la-Neuve : De Boeck, 2017. p. 97-108.

ARNOLD, D.; WAGNER, P. The influence of top-down expectations on the perception of syllable prominence. In: TUTORIAL AND RESEARCH WORKSHOP ON EXPERIMENTAL LINGUISTICS (ISCA), 2., Athens, Greece, 2008. *Proceedings...* Athens: University of Athens, 2008. p. 25-28,

ASTÉSANO, C. ; BERTRAND, R. Accentuation et niveaux de constituance en français : enjeux phonologiques et psycholinguistiques. *Langue Française*, Paris, v. 191, n. 3, p. 11-30, 2016. Doi: 10.3917/lf.191.0011

AVANZI, M. Note de recherche sur l'accentuation et le phrasé à la lumière des corpus du français. *Tranel*, Neuchâtel, Suisse, v. 58, p. 5-24, 2013.

CHRISTODOULIDES, G.; AVANZI, M. An Evaluation of Machine Learning Methods for Prominence Detection in French. In: ANNUAL CONFERENCE OF THE INTERNATIONAL SPEECH

COMMUNICATION ASSOCIATION, 15., 2014, Singapore. *Proceedings...* Singapore : International Speech Communication Association (ISCA), 2014. p. 116-119.

DI CRISTO, A. *Les musiques du français parlé*. Berlin : De Gruyter Mouton, 2016.

GARDE, P. *L'accent*. Paris : Presses Universitaires de France, 1968. (Collection SUP « Le linguiste », n. 5)

GARDE, P. *L'accent*. Paris : Lambert-Lucas, 2013.

GOLDMAN, J-P. ; AUCHLIN, A. ; ROEKHAUT, S. ; SIMON, A-C. ; AVANZI, M. Prominence perception and accent detection in French. A corpus-based account. *Language Science*, Elsevier, v. 39, p. 95-106, 2013.

GHITZA, O.; GREENBERG, S., On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, Basel, Suisse, v. 66, n. 1-2, p. 113-126, 2009.

LEHKA, I. ; LE GAC, D. Etude d'un marqueur prosodique de l'accent de banlieue. In : JOURNÉES D'ÉTUDES SUR LA PAROLE, XXV., Fès, Maroc, avril 2004. *Actes...* Fès, Maroc : Association Francophone de la Communication Parlée, 2004. Disponible sur : <http://www.afcp-parole.org/doc/Archives_JEP/2004_XXVe_JEP_Fes/actes/jep2004/Lehka-LeGac.pdf>. Accessed on January 12th, 2018.

LANGUE FRANÇAISE. *La prosodie du français : accentuation et phrasé*. Paris , v. 191, n. 3, 2016. 150 p. Doi: 10.3917/lf.191.0005

MARTIN, Ph. Analyse phonologique de la phrase française. *Linguistics*, v. 146, p. 3568, Fév. 1975.

MARTIN, Ph. La transcription des proéminences accentuelles : mission impossible ? *Bulletin PFC*, Toulouse, n. 6, p. 81-87, Sept. 2006.

MARTIN, Ph. Spontaneous speech corpus data validates prosodic constraints. In: INTERNATIONAL CONFERENCE ON SPEECH PROSODY, 6., 2012, Shangai, China. *Proceedings...* Shangai: Tongji University Press, 2014. p. 525-529.

MARTIN, Ph. *The Structure of Spoken Language. Intonation in Romance*. Cambridge: Cambridge University Press, 2015.

MARTIN, Ph. *Intonation, structure prosodique et ondes cérébrales*. London: ISTE Editions, 2018.

MEIGRET, L. *Le tretté de la grammaire françoéze*. Paris : C. Wechel, 1550. Disponible sur : <<http://gallica.bnf.fr/ark:/12148/bpt6k507854/fl.image>>. Accessed on January 12th, 2018.

MERTENS, P.; SIMON, A-C. Towards automatic detection of prosodic boundaries in spoken French. In: PROSODY-DISOURSE INTERFACE CONFERENCE (IDP-2013), 2013, Leuven. *Proceedings...* Leuven: KU Leuven, 2013. p. 81-87.

METTOUCHI, A.; LACHERET-DUJOUR, A.; SILBER-VAROD, V.; IZRE, Shlomo El. Only Prosody ? Perception of speech segmentation in Kabyle and Hebrew. *Cahiers de Linguistique Française*, Genève, v. 28, p. 207-218, 2007.

MICHELAS, A.; FRAUENFELDER, U. H.; SCHÖN, D.; DUFOUR, S. How deaf are French speakers to stress? *Journal of the Acoustical Society of America*, [s.l.], v. 139, n. 3, p. 1333-1342, 2016. Doi: <https://doi.org/10.1121/1.4944574>

ROSSI, M. Le seuil de glissando ou seuil de perception des variations tonales pour la parole. *Phonetica*, Aix-en-Provence, n. 23, p. 1-33, 1971. Doi:10.1159/000259328

SELKIRK, E. O. On prosodic structure and its relation to syntactic structure. In: FRETHEIM, T. (Ed.). *Nordic Prosody II*. Trondheim: TAPIR, 1978. p. 111-140.

VÄÄNÄNEN, V. *Introducción al latin vulgar*. Madrid: Editorial Gredos, 1995.

WIOLAND, F. *Les structures rythmiques du français*. Paris : Slatkine-Champine, 1985.



Prosodic Segmentation and Grammatical Relations: the Direct Object in Kabyle (Berber)

Segmentação prosódica e relações gramaticais: o objeto direto em kabyle (berbere)

Amina Mettouchi

École Pratique des Hautes Études, PSL, CNRS LLACAN, Paris / France

aminamettouchi@me.com

Abstract: The aim of the present paper is to provide evidence for the existence, in Kabyle (Berber), of the grammatical role ‘Direct Object’, and to define it using a non-aprioristic, empirical methodology. The definition, based on the analysis of corpus data, involves formal means pertaining to morphology, syntax and prosody. Prosodic segmentation is not only crucial for the definition of the category; it also serves as supporting evidence for the tightness of the relationship between verb and direct object.

Keywords: direct object; prosody; segmentation; Kabyle; Berber; grammatical relations.

Resumo: O objetivo deste artigo é oferecer evidências para a existência da função gramatical “Objeto Direto” em kabyle (berbere) e defini-lo utilizando uma metodologia empírica, não-apriorística. A definição, baseada na análise de dados de corpus, envolve meios formais pertinentes à morfologia, sintaxe e prosódia. A segmentação prosódica não é apenas crucial para a definição da categoria, mas também serve como evidência em favor da coesão da relação entre verbo e objeto direto.

Palavras-chave: objeto direto; prosódia; segmentação; kabyle; berbere; relações gramaticais.

Submitted on January 16th, 2018.

Accepted on May 12th, 2018.

Introduction

It is not so usual to associate prosody, especially prosodic segmentation, with the analysis of grammatical relations, but as previous work has shown (METTOUCHI, 2013, 2015, 2018 [2011]), considering prosodic cues as formal means with as much structuring potential as linear ordering or morphological marking actually allows the discovery of constructions in the domain of grammatical relations (and other domains too, such as information structure).

Within an empirical, and corpus-based approach, my purpose in this paper is to provide evidence for the existence, in Kabyle, of the grammatical role ‘Direct Object’, different from the semantic role ‘Referential undergoer’, and to define the Direct Object function in a non-aprioristic and language-internal perspective. I show that prosodic boundaries are crucial for the definition of Direct Object function in Kabyle, and that prosodic disfluencies provide evidence for the fact that the verb and its direct object form a constituent.

The paper first provides background information about Kabyle, prosodic units, and grammatical relations. The encoding of grammatical and semantic relations on bound pronouns is then analyzed, and I show that so-called ‘direct object pronouns’ in fact code ‘referential undergoer role’, a function different from direct object. In a third part, noun phrases coreferent to those bound pronouns are characterized, taking into account prosodic boundaries, and I show that their function is within the domain of referent activation, not grammatical relations. In a fourth part, the only noun phrase not coreferent with a bound pronoun, the direct object, is formally defined using syntactic, morphological and prosodic criteria. In a fifth part, proof is given of the tight relationship between verb and direct object, through the analysis of disfluencies and F0 contour.

1. Preliminaries

1.1 Kabyle

Berber languages are spoken in northern Africa, in a zone delimited by the Atlantic Ocean to the West, the Mediterranean to the North, the oasis of Siwa (Egypt) to the East, and the southern borders of Mali and Niger to the South. Those languages constitute a family within the Afroasiatic phylum. Well-known members of the family are, among others, Kabyle (spoken in northern Algeria), Tashelhiyt (Shilha) (spoken

in southern Morocco), and Tamashek and Tahaggart (also called Tuareg), spoken in southern Sahara.

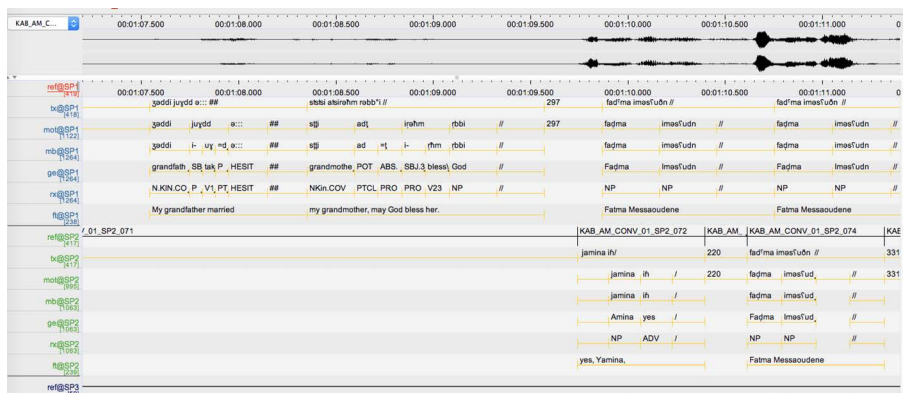
Kabyle has about four million speakers in the north of Algeria. The variety investigated in this paper is a Western one, spoken in the village of Ait Ikhlef, close to the town of Bouzeguene. I collected all the data on fieldwork between 2007 and 2011.

In Kabyle, as in all Berber languages, a minimal predication consists either of a verb and its bound personal pronoun, or of a non-verbal predicate. In this paper I focus on verbal predicates. In addition to this core, the clause may contain noun phrases, and prepositional phrases, as well as adverbs. Within noun phrases, modifiers follow the modified constituent. The language has two genders and two numbers, marked on adjectives, on nouns, and on pronominal affixes and clitics hosted by verbs, nouns and prepositions. It also has two states, marked on nouns.

1.2 Prosodic units and corpus

The corpus on which the study is based comprises, but is not limited to, one hour¹ of transcribed, segmented, annotated and translated narratives and conversations, collected in the field in Kabylie (Algeria) between 2007 & 2011. Examples in this paper are taken from the one-hour corpus.

FIGURE 1 – Layout of the Kabyle Corpus



¹ The Kabyle corpus is accessible and searchable online, at <<http://dx.doi.org/10.1075/sc1.68>. website>.

As shown in Figure 1, morphosyntactic annotation² is displayed on two tiers, “ge” and “rx”, allowing the automatic retrieval of complex queries based on forms. In the case of grammatical relations, only one-to-one form-function mappings were annotated: distinct and dedicated pronominal paradigms were given the label corresponding to their function (see below for expanded discussion on pronouns), but since nouns are only marked for gender, number and state, and since state does not code grammatical role (see below and METTOUCHI, 2018 [2011]), no noun was annotated as subject, direct or indirect object. The question of whether grammatical relations are coded for nouns, through constructions, was left open for investigations, which have been conducted using a query language based on regular expressions.³

An intonation unit is a segment of speech that has a coherent intonation contour (CHAFE, 1994), and is delimited by its boundaries (CRUTTENDEN, 1997), which bear a ‘boundary tone’ (PIERREHUMBERT; HIRSCHBERG, 1990). In Kabyle, Intonation Units are marked by one or more of the following cues:⁴

Main external cues

- (1) final lengthening; (2) initial rush; (3) pitch reset; (4) pause; (5) creaky voice.

Main internal cues

- (1) declination; (2) tonal parallelism, or isotony.

² The following abbreviations are used: ABS absolute state; ABSV absolute pronominal paradigm; ANN annexed state; AOR aorist; ASSOC associative; CAUS causative; CNS shared reference demonstrative; COL collective; COM comitative; COP copula; DAT dative; EXNEG existential negative; F feminine; GEN genitive; HESIT hesitation; IDP independent pronoun; IPFV imperfective; KIN kinship pronominal paradigm; M masculine; NEG negation; PFV perfective; PL plural; POS positive; POSS possessive pronominal paradigm; PREP prepositional pronominal paradigm; PROX proximal; SBJ subject pronominal paradigm; REAL realis; REL relator; RELSBJ subject relativization circumfix; SG singular; VOC vocative. A list of glosses with definitions, explanations and references can be found on http://corpafroas.huma-num.fr/Archives/KAB/PDF/KAB_AM_ALISTOFGLOSSES.PDF

³ For the syntax of queries using regular expressions, see http://llacan.vjf.cnrs.fr/fichiers/manuels/ELAN/ELAN-Corpa_Search.pdf

⁴ See (IZRE’EL; METTOUCHI, 2015) for more information on the segmentation of the CorpAfroAs corpus of spoken Afroasiatic languages, of which the Kabyle corpus is a part.

The data were segmented into intonation units⁵ on the basis of native speaker perception, and acoustic control with Praat.⁶ Two native speakers were first made to understand that what was asked of them concerned the melodic and rhythmical contour of the unit, not its lexical, grammatical or pragmatic contents. Then the recording was played using Praat, and they were asked to tell where they would insert boundaries in the flow of speech; they indicated that by a beat of the hand on the table. For each beat, the annotator inserted a boundary marker in the Praat textgrid corresponding to the sound file. The units thus delimited were additionally controlled with Praat whenever there was disagreement between the two native speakers. Later, the annotator added units for silent pauses over 200 ms (the number inside indicates duration of the pause in milliseconds), and for breath intakes (coded as BI, followed by the duration of the intake in milliseconds). All examples used for this study were systematically re-controlled with Praat.

The total number of non-pausal intonation units for the whole one-hour corpus is 2671. With breath intakes and silent pauses, the total number of units⁷ is 3974.

Intonation Units are usually considered as either linked to the domain of cognition (CHAFE, 1994) or pragmatics (CRESTI; MONEGLIA, 2005) in some approaches, or as the projection of clause structure (syntactic level) onto the prosodic level (SCHEER, 2011; SELKIRK, 2009; VOGEL, 2006), in other approaches.

In the first type of approaches, intonation units are seen as encapsulating an 'idea' (CHAFE, 1994), or a 'speech-act' (CRESTI; MONEGLIA, 2005). However, the existence of intonation units (formally defined by prosodic cues only) that are not pragmatically autonomous, such as in the following example, points to the fact that there is no necessary mapping between idea/speech-act and intonation unit.

⁵ Annotated as /: non-terminal boundary; //: terminal boundary.

⁶ <http://www.fon.hum.uva.nl/praat/>

⁷ Each unit has been numbered following a precise methodology : ISO code of the language, initials of the author, genre (NARR is for narrative, CONV for conversation), number of the file, number of the unit. Thus, all examples are easily found in the corpus.

- (1) *urdəzwidʒəy / BI_363 / alamma θəksədd / fatima θuhrʕiθ / 326 / ayrum əgðəkkwan //*
 ur=dd zwiḡ-y / BI-363/ alamma
 NEG=PROX marry\NEGPFV-SBJ.1SG / BI_363 / until
 PTCL=PTCL V23-PRO / BI_363 / CONJ
 t-əks=dd / faṭima tuhrjʔt / 326 /
 SBJ.3SG.F-take_away\PFV=PROX / Faṭima clever / 326 /
 PRO-V23=PTCL / NP ADJ / 326 /
 ayrum g udəkkʰan //
 bread\ABS.SG.M LOC shelf\ANN.SG.M //
 N.OV PREP N.OV //
 I won't marry / ... / until she grabs / clever Fatima / ... / the bread on the shelf //
 (I won't marry until Clever Fatima grabs the bread on the shelf)
 (KAB_AM_NARR_01_0086-91)



As for the second type of approaches, as shown by Tao (1996) and Ross (2011) among others, there is no one-to-one mapping between clause and intonation unit: in Kayarldid and Dalabon (ROSS, 2011), intonation units are more commonly found to comprise part of a clause, and discourse factors override grammatical phrasing (there are multi-verb intonation units, as well as single NP intonation units). As for Mandarin, Tao (1996) notes the high number of NP units, as well as elliptical clausal intonation units.

However, the fact that intonation units are not projections of syntax onto prosody doesn't mean that there is no link between intonation unit and clausal structure. Indeed, all intonation units that contain a verb in Kabyle necessarily contain a minimal clause, since no verb can appear without its obligatory subject affix. A typology of such units has been proposed in Mettouchi (2013, 2015, 2018 [2011]), on the basis of the presence, ordering and morphology of noun phrases inside or at the periphery of the prosodic group containing the verb.

In that typology, the prosodic group containing the verb is defined as a unit whose left and right boundaries, marked by such cues as final lengthening, initial rush, pitch reset, pause, and/or creaky voice, enclose a verb. The types found in Kabyle spontaneous discourse (narratives and conversation) are the following (METTOUCHI, 2013, 2015, 2018 [2011]):⁸ (a) $[V_{\text{sbj}} (N_{\text{ABS}})]$; (b) $[V_{\text{sbj}} N_{\text{ANN}} (N_{\text{ABS}})]$; (c) $[N_{\text{ABS}} V_{\text{sbj}} (N)]$; (d) $N_{\text{ABS}} [V_{\text{sbj}} (N) (N)]$, and (e) $[V_{\text{sbj}} (N) (N)] N_{\text{ANN}}$, where prosodic

⁸ V_{sbj} represents the verb and its obligatory subject affix (other clitics may also attach to the verb), N_{ann} represents a noun in the annexed state, and N_{abs} a noun in the absolute state.

boundaries are represented by square brackets. Illustrative examples are given below, and the reader is referred to the abovementioned publications for contextualized examples taken from spontaneous corpora.

- (a) [j-ʃʃa (ayr^um)]SBJ3.SG.M-eat\PFV (bread\ABS.SG.M)
‘He ate (bread)’
- (b) [j-ʃʃa wqɪʃ (ayr^um)]SBJ3.SG.M-eat\PFV child\ANN.SG.M (bread\ABS.SG.M)]
‘The boy ate (bread)’
- (c) [aqɪʃ j-ʃʃa ayr^um]child\ABS.SG.M SBJ3.SG.M-eat\PFV bread\ABS.SG.M
‘The boy ate bread’
- (d) aqɪʃ [j-ʃʃa (ayr^um)]child\ABS.SG.M /SBJ3.SG.M-eat\PFV (bread\ABS.SG.M)
‘The boy, he ate (bread)’
- (e) [j-ʃʃa (ayr^um)] wqɪʃ-nni [SBJ3.SG.M-eat\PFV (bread\ABS.SG.M)/ child\ANN.SG.M-CNS]
‘He ate (bread), that boy’

Those types have information structure and referent activation functions, and the grammatical role of nouns is not systematically coded by the construction (METTOUCHI, 2018 [2011]). The information structure function of such constructions as (a) is (sub-) topic continuation: the protagonist is the same, and the narrative is carried forward; (b) introduces a new episode in a narrative or a new subtopic in a conversation; (c) builds a background for further developments, recapitulating a salient preceding situation, so that the listener grasps the whole situation and its importance for the current discourse; (d) marks a shift in perspective or contrast with previous expectations; and (e) reactivates a referent that had lost its active or semi-active status (METTOUCHI, 2015, 2018 [2011]).

As mentioned above, not all nouns are transparently coded (i.e. formally recognizable, vs. retrieved by inference only) for grammatical role in Kabyle. Detailed evidence is given in Mettouchi (2018 [2011]) in support of that claim.

1.3 Grammatical relations

My approach does not consider as a given the fact that in Kabyle, grammatical relations are encoded on all nominal/pronominals. Indeed, as shown in Mettouchi (2013, 2018 [2011]), unless preceded by a preposition, only some nominals, those inside the prosodic group of the verb, may be attributed a grammatical role. Nominals belonging to

the sentence, but situated before or after the prosodic boundaries of the prosodic group of the verb can be coreferent to a bound pronoun that has a given grammatical or semantic role, but they do not, either through morphology or construction, encode such roles. Their function is more centrally in the domain of information structure and referent activation, as mentioned in the preceding part (see METTOUCHI, 2015, 2018 [2011]) for extended and commented examples from my corpus).

In that respect, this study differs from works that take the existence of the category ‘direct object’ as not needing to be established within a specific language, nor defined in a more formal way than in Matthews (2007) for instance:

direct object (DO) An *object traditionally seen as identifying someone or something directly involved in an action or process: e.g. *my books* in *I might leave my books to the library*, where it is distinguished from the *indirect object *to the library*. Hence, in particular, the object typically next to the verb in English, one marked by the accusative case in German, and so on. (MATTHEWS, 2007, p. 106)

object (O) 1. An element in the basic sentence construction of a language such as English which characteristically represents someone or something, other than that represented by the *subject (1), that is involved in an action, process, etc. referred to. E.g. *him* in *I met him*; both *her* and *afflower* (respectively the *indirect object and the *direct object) in *I will give her afflower*; also, on the assumption that it is syntactically the same element, *that I did* in *I said that I did*. 2. An element seen as standing in a similar relation to a preposition: e.g. *Washington* in *from Washington*. 3. Any element, in any type of language, which characteristically includes the semantic role of *patient. Cf subject (3): thus, in typological studies, a language may be classified as an *SVO language simply because that is the commonest order, in texts, of agent, verb, and patient. (MATTHEWS, 2007, p. 272)

patient (P) 1. Noun phrase or the equivalent that identifies an individual etc. undergoing some process or targeted by some action. E.g. *the house* is a patient in *I painted the house*; *Mary* in *I kissed Mary*. 2. Thence of a syntactic role which is characteristically that of a patient. E.g. a **direct object** in English tends to be a patient, especially a patient rather than an *agent.

Therefore **direct objects** and elements in other languages which are in this respect equivalent to them may be called, in general, patients.

The sense is that of Latin *patiens*, ‘suffering’ or ‘undergoing’. Abbreviated to P especially in cross-linguistic studies, where opp. A for *agent (2); also opp. S (3). (MATTHEWS, 2007, p. 290)

My approach also differs from studies that, having taken the category ‘direct object’ for granted, and having either selected typical examples from corpora, or having created sentences for reading experiments, provide findings about ‘the prosody of direct objects’.

While I acknowledge the importance and relevance of those studies, my perspective is different in that it includes prosodic forms in the very definition of the category in Kabyle: there is no ‘prosody of direct objects’, but rather, a construction involving syntactic, morphological and prosodic forms which (a) encodes the ‘direct object’ function, and (b) translated into an automatic query, allows the retrieval of all and only the direct objects in a spoken corpus of Kabyle, non-aprioristically annotated according to forms.

2. Bound pronouns and their roles

While nominals are most of the time absent, bound pronouns are noticeable and frequent in Kabyle. The language has several pronominal paradigms (METTOUCHI, 2017, p. 10-11). Among those, some are hosted by the verb: the subject affix, the absolutive clitic, and the dative (indirectly affected argument) clitic.

Subject pronouns are affixes (their position relative to the verb is fixed), and only appear with verbs; dative pronouns are clitics (they undergo climbing in contexts of negation, relativization, or irrealis mood). This is also true for absolutive pronouns, which, additionally, are also hosted by some non-verbal predicates (they are their sole argument).

2.1 Subject affixes and dative (indirectly affected argument) clitics

Subject affixes code various participant roles, among them sole argument of intransitive verbal constructions (2), affecting argument of active transitive constructions (3), and affected argument of passive transitive constructions (4).

(2) atsali arθkana /

ad	t-ali	ar	tkanna	/
POT	SBJ.3SG.F -go_up\AOR	to	attic\ANN.SG.F	/
PTCL	PRO-V14	PREP	N.OV	/

'she would go up to the attic'
(KAB_AM_NARR_01_0862)



(3) θssuliθid /

t-ssuli=t-dd	/
SBJ.3SG.F -go_up\CAUS.PFV=ABSV.3SG.M=PROX /	
PRO-V14=PRO=PTCL /	

'she pulled him up'
(KAB_AM_NARR_01_0968)



(4) aθtʃwəθjənt /

ad	tʃwəθθ-nt	/
POT	eat\PASS.AOR- SBJ.3PL.F /	
PTCL	V13%-PRO	/

'(the little girls were) to be eaten alive'
(KAB_AM_NARR_01_0710)



Dative clitics code the indirectly affected argument: addressee, recipient as in (5), positively or negatively affected participant as in (6)...

(5) θəfkajasəntətʃ /

t-fka=asnt=t	/
SBJ.3SG.F -give\PFV= DAT.3PL.F =ABSV.3SG.M /	
PRO-V13%=PRO=PRO	/

'she gave it to them (her sisters)'
(KAB_AM_NARR_01_0537)



(6) <ça fait> θəmmuθas θəqʃiʃθ iZafiwa θaʃlits //

<ça fait>	t-mmut=as	təqʃiʃt	i
it_is	SBJ.3SG.F -die\PFV= DAT.3SG	girl\ANN.SG.F	DAT
CSW.FRA	PRO-V24=PRO	N.OV	DEMPRO

Zafiwa	Taʃliʃ
Zafiwa	daughter_of_Ali
N.P.	N.P.

'you were saying she lost a daughter (lit. 'a girl died on her'), Zahwa Taʃliʃ ?'

θəmmuθas θmənzuθ //


t-mmut=as	tmənzut	//
SBJ.3SG.F -die\PFV= DAT.3SG	elder\ANN.SG.F	//
PRO-V24=PRO	N.OV	//

'Her eldest daughter died (on her)'

(KAB_AM_CONV_01_SP3_31 & SP1_276)




Moreover, not any type of undergoer is thus encoded: the participant has to be referential, it cannot be non-referential or non-existent, as shown by (10) and (11).

- (10) *addənsəw / ulaf ipajasən / ulaf əlqaʃa*
- | | | | | | |
|-----------|-------------------|---|---------|-------------------|---|
| ad=dd | n-səw | / | ulaf | ipajasn | / |
| POT=PROX | SBJ.1PL-drink\AOR | / | NEG.EXS | mattress\ABS.PL.M | / |
| PTCL=PTCL | PRO-V23% | / | PRED | N.OV | / |
-
- | | | | |
|---------|-----------------|---|--|
| ulaf | lqaʃa | / |  |
| NEG.EXS | ground\ABS.SG.F | / | |
| PRED | N.COV | / | |
- 'we would drink, there were no mattresses, there was no proper ground'*
(KAB_AM_NARR_03_0239-242)

Indeed, negative existential predication is coded by *ulaf*, without any pronoun, possibly followed by a noun in the absolute state, as in (10), whereas *ulaf* hosting an absolutive bound pronoun (as in (8)) cannot express absence of a referent, it necessarily means that the referent exists but not at that location.

Referentiality of the absolutive pronoun is also a property of verbal predications: an absolutive pronoun cannot co-refer with an abstract or non-referential noun, as shown by the ungrammaticality of examples (11') and (11''), constructed from the original formulation in (11).

- (11) *θəsʃa lhərʃma /*
- | | | | |
|-----------------------|--------------------------|---|--|
| t-sʃa | lhərma | / |  |
| SBJ.3SG.F-possess\PFV | good_reputation\ABS.SG.F | / | |
| PRO-V13% | N.COV | / | |
- 'she had good reputation'*
(KAB_AM_NARR_03_0567)
- (11') *lhərma, t-sʃa=tt
good_reputation\ABS.SG.F SBJ.3SG.F-possess\PFV=ABSV.3SG.F
N.COV PRO-V13%=PRO
*good reputation, she had it.
- (11'') *t-sʃa=tt, lhərma
SBJ.3SG.F-possess\PFV=ABSV.3SG.F good_reputation\ANN.SG.F
PRO-V13%=PRO N.COV
*she had it, good reputation.

Those characteristics lead me to define absolutive pronouns in Kabyle as coding the role of referential undergoer.

Pronominal paradigms hosted by verbs are therefore not homogeneous in terms of categories: whereas the subject affix clearly codes grammatical role, the absolutive and the dative bound pronouns code semantic roles in Kabyle.

3. Coreferent nominal

As is the case for all bound pronouns in Kabyle, the referent of absolutive pronouns can be expanded by a coreferent nominal. Whereas pronouns come in various paradigms, nouns must be either in the absolute or in the annexed state. This binary morphological marking, covert in the case of borrowings and for some classes of nouns with a special phonological structure, is marked differently depending on the gender and the number of the noun (table 1).

TABLE 1 – Gender, Number and State in Kabyle

	Masculine		Feminine	
	Singular	Plural	Singular	Plural
Absolute	a-mȳar	i-mȳar-n	t-a-mȳar-t	t-i-mȳar-in
Annexed	w-mȳar	j-mȳar-n	t-mȳar-t	t-mȳar-in

(root *mȳar*, ‘old person’)

The state distinction plays a structural role in the language. It is the backbone of the whole grammatical system of Kabyle and is functional at the level of the phrase as well as at the level of the clause and the sentence (METTOUCHI; FRAJZYNGIER, 2013; METTOUCHI, 2014)

In Kabyle, the function of the annexed state is to “provide the value (in the logical sense) for the variable of the function grammaticalized in a preceding constituent”⁹ (METTOUCHI; FRAJZYNGIER, 2013, p. 2),

⁹ “A grammaticalized function is a function that is represented by a morpheme, which may be affixal (bound pronouns, gender-number markers) or non-affixal (prepositions, relational morphemes). A function is grammaticalized when it is coded by some grammatical marker.” (METTOUCHI; FRAJZYNGIER, 2013, p. 2)

while the absolute state “is the default form of the noun and does not carry any specific function.” (METTOUCHI; FRAJZYNGIER, 2013, p.2).

Nouns in the annexed state always follow the marker for whose function they are a variable. Therefore, a noun in the annexed state cannot be the first element of any structure in Kabyle. Nouns in the annexed state can be complement of prepositions, of relational nouns, they can be coreferent to a pronoun bound to a verb or a noun... Nouns in the absolute state are not constrained in position or function; in a binary system where nouns must be in either the annexed or the absolute state, their contexts of occurrence are in complementary distribution with the contexts of the annexed state, they are the default member of the opposition. This does not prevent them from being part of constructions which are themselves functional: ‘verb followed by noun in the absolute state’ is a construction with a function no less marked than ‘verb followed by noun in the annexed state’.

3.1 Computing coreferentiality

Coreferent nouns are in the absolute state when they precede the functional element with which they are coreferent, here the subject pronoun (12), and in the annexed state when they follow it (13).

(12) *argaz ađir^{uh} ađjawi θajuja ađiçrəz /*

argaz	ad	i-ruh	ad	j-awi
man\ABS.SG.M	POT	SBJ.3SG.M-go\AOR	POT	SBJ.3SG.M-bring\AOR
N.OV		PTCL PRO-V24		PTCL PRO-V14

tajuga	ad	i-kərz /
pair_of_oxen\ABS.SG.F	POT	SBJ.3SG.M-plough\AOR /
N.OV		PTCL PRO-V23.LAB /

The husband would go and bring a pair of oxen to plough,

(KAB_AM_NARR_03_0096)



(13) *nəy ma issuθidd wərgazis /*

nəy ma	i-ssutr=as=dd	wərgaz-is /
or if	SBJ.3SG.M-request\PFV.CAUS=DAT.3SG.M=PROX	man\ANN.SG.M-
CONJ CONJ	PRO-V24=PRO=PTCL	POSS.3SG /
		N.OV-PRO /

or when her huband requests something,

(KAB_AM_NARR_03_1125)




Coreference is computed on the basis of identity of gender and number between pronoun and noun. In some cases of ambiguity (e.g. same number and gender on the subject and absolutive bound pronouns), establishment of coreference also relies on probabilistic inferences.


3.2 Nouns coreferent to absolutive pronouns

No noun in the annexed state coreferent to an absolutive pronoun appears within the prosodic group of the verb in my data. Nouns in the annexed state within the prosodic group of the verb are all coreferent with subject affixes, as in example (13). And nouns in the annexed state coreferent with absolutive pronouns are outside the prosodic group of the verb, always after a prosodic boundary, as in (14) and (15):

- (14) *tufa ðamfi/buðrar // 423 iθizəðγən / wəχχamni //*
 t-ufa d amfiʃ n wəðrar // 423
 SBJ.3SG.F-find\PFV COP cat\ABS.SG.M GEN mountain\ANN.SG.M //
 PRO-V13% PRED N.OV PREP N.OV //

 i=t i-zdəγ-n / wəχχam-nni //
 REL.REAL=**ABSV.3SG.M** RELSBJ.POS-dwell\PFV-RELSBJ.POS /**house\ANN.SG.M-CNS**//
 DEMPRO=PRO CIRC1-V23-CIRC2 / N.OV-DEM //
 ‘she found it was the Mountain Cat who inhabited it,the house’
 (KAB_AM_NARR_01_0414-417) 


- (15) *aytniddəfk sətsi antnəff / BI_404 ihβuβənənni /*
 ad=ay=tn=dd t-əfk sətʃi
 POT=DAT.1PL=**ABSV.3PL.M**=PROX SBJ.3SG.F-give\AOR grandmother\
 ANN.SG.F
 PTCL=PRO=PRO=PTCL PRO-V13% N.KIN.COV

 ad=tn n-čč / jəhbubən-nni /
 POT=**ABSV.3PL.M** SBJ.1PL-eat\AOR / **dried_figs\ANN.PL.M-CNS** /
 PTCL=PRO PRO-V13% / N.OV-DEM /
 ‘my grandma would give them to us so that we would eat them, those dried figs’
 (KAB_AM_NARR_03_0353-0355) 

Prosodic boundaries are therefore crucial in the interpretation of relations between the participants in a state of affairs.

It is not, however, true that any noun in the annexed state after the prosodic group of the verb necessarily corefers to an absolutive

pronoun: such nouns can be coreferent to other types of bound pronouns, including those hosted by nouns. Here is an instance of coreference with the subject affix:

- (16) *dəməddaʃ# #¹⁰ θa# θasumtanni itssummuθ akkən / wəmʃifənni /*
 t-ddəm=dd aʃ# ##
 SBJ.3SG.F-grab\PFV=PROX FS# ##
 PRO-V23=PTCL FS# ##
- ta# tasumta-nni i-ʃsummut akk-ən /
 FS# pillow\ABS.SG.F-CNS **SBJ.3SG.M-use_as_pillow**IPFV thus-DIST /
 FS# N.OV-DEM PRO-V24.PFX.APHO ADV-AFFX /
- wəmʃif**-nni /
cat\ANN.SG.M-CNS /
 N.OV-DEM /
- 'she took the pillow on which he slept, the cat'*
 (KAB_AM_NARR_01_0445-0448)
- 

One cannot therefore consider that a noun in the annexed state following the prosodic group of the verb is a direct object (or more generally, that it has a grammatical role, given that it can corefer with several types of pronouns). As mentioned in part 1.2., its role is within the domain of referent activation – the noun is used to reactivate a referent that had lost its active or semi-active status (METTOUCHI, 2018 [2011], p. 273).

The same is true for nouns preceding the prosodic group of the verb, and co-referent with absolutive pronouns. They are in the absolute state (as are all nouns preceding the verb), and encode contrastive comments (METTOUCHI, 2018 [2011], p. 272), regardless of their coreferent pronoun (e.g. the subject affix in example (17)).

¹⁰ A single crosshatch # at the end of a sequence of syllables indicates a truncated word. A double crosshatch following a series of words indicates a truncated intonation unit (in general due to disfluencies, but also to interruptions in conversations). See (IZRE'EL; METTOUCHI, 2015) for the notion of abandoned intonation unit.

(17) *asənduqagi / atʰtʰawiðʰ ijəmmak //*

asənduq-agi /
chest\ABS.SG.M-PROX1 /
 N.OV-AFFX /



ad=t t-awi-d i jəmma-k //
 POT=**ABSV.3SG.M** SBJ2-bring\AOR-SBJ.2SG DAT mother\ANN.SG.F-
 KIN.2SG.M//
 PTCL=PRO CIRC1-V14-CIRC2 DEMPRO N.KIN.COV-PRO //
 'this box, you will take it to your mother'
 (KAB_AM_NARR_02_783-784)

A noun may also appear before the verb within the prosodic group of the verb, and be coreferent to an absolutive pronoun as in (18) below. Grammatical and semantic relations are marked by the bound pronouns, nouns are referential expansions of those pronouns, and the structure recapitulates the preceding events and situation in a condensed way, as a synthetic explanatory comment on the preceding discourse (METTOUCHI, 2015, p. 130).

(18) *azdduznni jʰsawiθuβəhri /*





azdduz-nni i-tʰawi=t ubəhri /
big_stick\ABS.SG.M-CNS SBJ.3SG.M-bring\IPFV=**ABSV.3SG.M** wind\ANN.
 SG.M /
 N.OV-DEM PRO-V14.PFX=PRO N.OV /
 'the wind moved the stick'
 (KAB_AM_NARR_01_0756)



In sum, in Kabyle, grammatical and semantic relations are coded by pronouns, and coreferent nouns are involved at other levels of speech organization: reference, referent activation, information structure.

4. Direct objects

Only one noun can appear within the prosodic group of the verb without being coreferent to a bound pronoun. It is in the absolute state, it follows the verb (which itself necessarily bears a subject affix), either immediately, or separated from it by an adverb, a postverbal negator, and/or a noun in the annexed state. This characterization I consider to be the formal definition of direct objects in Kabyle.

- (19) *jəddməttatsəffaht* /
 i-ddəm=dd **taʔəffaht** /
 SBJ.3SG.M-grab\PFV=PROX **apple**\ABS.SG.F /
 PRO-V23=PTCL N.OV / 
 'he took an apple'
 (KAB_AM_NARR_02_029)
- (20) *θətʃlaʃakka amʃijənni* //
 t-ʃlaʃi akk-a **amʃij-nni** //
 SBJ.3SG.F-address\IPFV thus-PROXa **cat**\ABS.SG.M-CNS //
 PRO V14.PFX ADV DEMPRO N.OV DEM // 
 'she addressed the cat'
 (KAB_AM_NARR_01_0597)
- (21) *innajas zʃran wajθma θazʃarʃbiθ arjəmmanuza* /
 i-nna=as zra-n wajtma **taʔərbit**
 SBJ.3SG.M-say\PFV=DAT.3SG see\PFV-SBJ.3PL.M brother\ANN.PL.M **carpet**\
ABS.SG.F
 PRO-V13%=PRO V13%-AFFX N.KIN.OV N.OV
 ar jəmma Nuʒa / 
 to mother\ANN.SG.F Nuʒa /
 PREP N.KIN.COV N.P /
 'he told him that his brothers had seen a carpet at Jemma Nuja's place'
 (KAB_AM_NARR_02_505)
- The noun refers to an undergoer, and can be abstract or concrete, referential or non-referential, effected or affected.
- (22) *unxəddəmarak° tʃumatifagi tətəwatʃ itʃujt* //
 ur n- xəddm ara ak° **tʃumatif-agi**
 NEG SBJ.1PL-make\IPFV POSTNEG all **tomato**\ABS.COL-PROXB
 PTCL PRO-V23 N.INDF ADV N.COVS-DEM
 n tətəwatʃ i tʃujt // 
 GEN can\ANN.SG.F LOC pot\ANN.SG.F //
 PREP N.OV PREP N.OV //
 'we didn't put tomato concentrate in the pot'
 (KAB_AM_NARR_03_0793)
- (23) *θəsʃa lhərʃma* /
 t- sʃa **lhərʃma** /
 SBJ.3SG.F-possess\PFV **good_reputation**\ABS.SG.F /
 PRO-V13% N.COVS /
 'she had good reputation'
 (KAB_AM_NARR_03_0567)

It is possible to automatically retrieve those nouns in the corpus by launching the query: ‘inside the prosodic group of the verb, look for a noun in the absolute state immediately following the verb or following <the verb followed by a noun in the annexed state (ANN in ge)> or following <the verb followed by an adverb (ADV in rx)> or following <the verb followed by a postverbal negator (POSTNEG in ge)>’.

Adverbs belong to a closed class and therefore are computed as such by the speaker or listener. The same is true for the postverbal element of negation, *ara*, of nominal origin, but grammaticalized as NEG2: its grammatical status is clear to the speaker or listener. Their intercalation between the verb and the noun in the absolute state pose no threat to interpretation.

Through the delimitation of a unit (the prosodic group of the verb) inside which the grammatical role ‘Direct Object’ can be transparently computed using forms and not probabilistic inferences, prosody plays an important role in the treatment of grammatical information in Kabyle, and this is also shown by sequences that apparently constitute counter-examples to my claims, but are actually evidence supporting them.

5. Prosodic disfluencies and constituency

Sometimes indeed, a noun in the absolute state appears after the prosodic boundary of the prosodic group of the verb, and this generally signals that a new clause is beginning:

- (24) *qqimən a?amina qqimən / ajθmasnak° ylin /*
- | | | | | |
|------------------------------------|------|-------|----------------------|----------------------|
| qqim -n | a | Amina | qqim-n | / |
| stay\PFV-SBJ.3PL.M | VOC | Amina | stay\PFV-SBJ.3PL.M / | |
| V24-PRO | PTCL | NP | V24-PRO | / |
| ajtma-tsn | | | ak° | yli-n / |
| brother \ABS.PL.M-KIN.3PL.M | all | | | fall\PFV SBJ.3PL.M / |
| N.KIN.OV-PRO | | | ADV | V24-PRO / |
- ‘They stayed Amina, they stayed, and his brothers all fell asleep’
 (KAB_AM_NARR_02_Midget_340-341)



But in some cases, the noun clearly belongs to the current clause:

- (25) ə:: / nəkki ffiydd ziqnni / θəssəhfəð^ɨiji jəmmaʃbb^oað^s / θəssəhfəð^ɨiji aɣrum /
 θəssəhfəð^ɨiji ləsfənz / θəssəhfəð^ɨiji ə:: / BI_412 a:: səksu /
 ə:: / nəkki ffi-γ=dd zik-nni /
 HESIT / IDP.1SG remember\PFV-SBJ.1SG=PROX long_ago-CNS /
 HESIT / PRO V13%-PRO=PTCL ADV-DEM /
- t-ssəhfəð=iji jəmma afwwað /
 SBJ.3SG.F-learn\CAUS.PFV=DAT.1SG mother\ANN.SG.F pancake_soup\ABS.
 SG.M /
 PRO-V24=PRO N.KIN.COV N.OV /
- t-ssəhfəð=iji aɣrum /
 SBJ.3SG.F-learn\CAUS.PFV=DAT.1SG bread\ABS.SG.M /
 PRO V24 PRO N.OV /
- t-ssəhfəð=iji lsfənz /
 SBJ.3SG.F-learn\CAUS.PFV=DAT.1SG doughnut\ABS.SG.M /
 PRO V24 PRO N.COV /
- t-ssəhfəð=iji ə:: /
 SBJ.3SG.F-learn\CAUS.PFV=DAT.1SG HESIT /
 PRO V24 PRO HESIT /
- a:: səksu /
 HESIT cuscus\ABS.SG.M /
 HESIT N.COV /
- 'ehm, I remember in the past, my mother taught me (how to cook) pancakes, she
 taught me (how to cook) bread, (how to cook) doughnuts, she taught me (how to cook)
 ehm, cuscus'*
- (KAB_AM_NARR_03_0192-0200)



It is important though, that this intuition be supported by formal criteria. Among the conditions listed in the preceding part, the fact that the noun is in the absolute state and the fact that there is no coreferent pronoun are met, but here the noun is not inside the prosodic group of the verb. Does that mean that one of the features of direct objects as I defined them is to be taken out of the definition? I argue that on the contrary, such examples in fact support my claim concerning the formal definition of the construction.

Indeed, we do not simply have a neat prosodic boundary separating the noun from the preceding verb. What we have, and that we can take into account thanks to a precise transcription of the spoken data, is a boundary that is so to say bridged by prosodic phenomena that are continuation cues: such nouns in the absolute state are systematically preceded, before the prosodic boundary, by:

- (a) a hesitation marker (this is the most frequent situation) (26)
- (b) a false start resulting in an interrupted IU, followed by a restart (27)
- (c) a rising boundary tone (28)

(26) *awnəʃkəy a::: / ifr^əaxa //*
 ad =wən əfk-y a::: / ifrax-a //
 POT=DAT.2PL.M give\AOR-SBJ.1SG **HESIT** / bird\ABS.PL.M-PROXa //
 PTCL=PRO V13%-PRO HESIT / N.OV-AFFX //
'and I'll give you ehm... those birds'
 (KAB_AM_NARR_02_176-177)



(27) *ixədmas θa:::# ## θabburθ duzzal /*
 i-xdəm=as ta:::# ##
 SBJ.3SG.M-make\PFV=DAT.3SG **FS:::# ##**
 PRO-V23=PRO FS ##
 tabburt d uzzal /
 door\ABS.SG.F COP iron\ABS.M.SG /
 N.OV PRED N.OV /
'he put on it an iron door'
 (KAB_AM_NARR_02_708-710)



(28) *asjini / hafama θəʃkawθəddak^o / θaδ^oaδəʃt taδ^oaδəʃθ //*
 ad=as j-ini /
 POT=DAT.3SG SBJ.3SG.M-say\AOR /
 PTCL=PRO PRO-V13% /
 hafama t- fka-wt=dd ak^o /
 only_when SBJ2-give\PFV-SBJ.HORT.2PL=PROX all /
 CONJ CIRC1-V13%-CIRC2=PTCL ADV /
 taɖaɖəʃt taɖaɖəʃt //
 finger\ABS.SG.F finger\ABS.SG.F //
 N.OV N.OV //
'he would say, not until you each give me, one of your fingers'
 (KAB_AM_NARR_02_171-173)



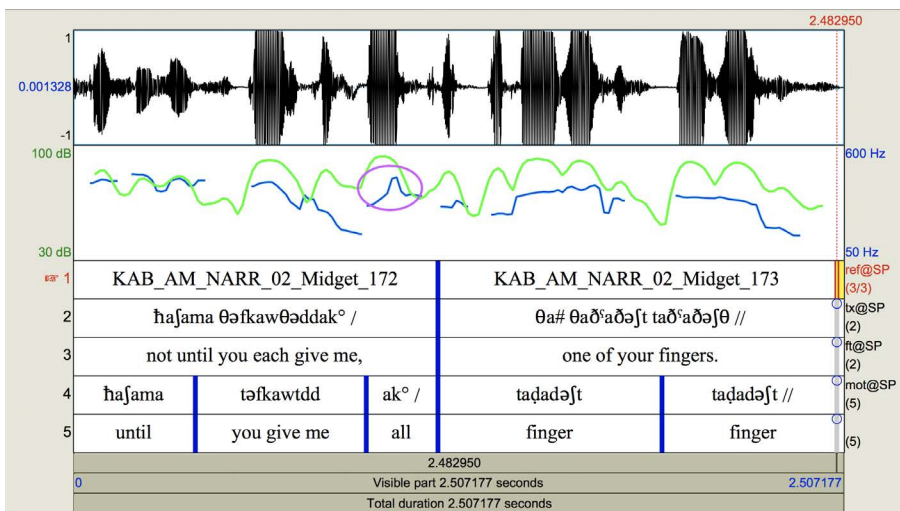
The first two phenomena are disfluencies, they might very well have ended up in an abandoned intonation unit, followed by a complete syntactic reformulation. But this never happens with direct objects in my data. On the contrary, disfluencies, some linked to planning issues and others to situational factors in the interaction, are systematically filled in by prosodic materials pertaining to continuation strategies, such as lengthening of a hesitation marker (itself a filler) or of a false start, and the sequence is immediately resumed in the form of the expected noun

in the absolute, the direct object: *ifrax-a* in (26), *tabburt* in (27). This can also be seen in Figure 1 and example (16).

This shows that in terms of cognitive processing, there is a strong relationship between the prosodic group of the verb and its stranded object.

The third phenomenon involves a continuative boundary tone, as shown in the Praat picture below on *ak°*, in example (28), with a value of 445 Hz:

FIGURE 2 – Praat acoustic analysis of example 28 (F0 and Intensity curves)



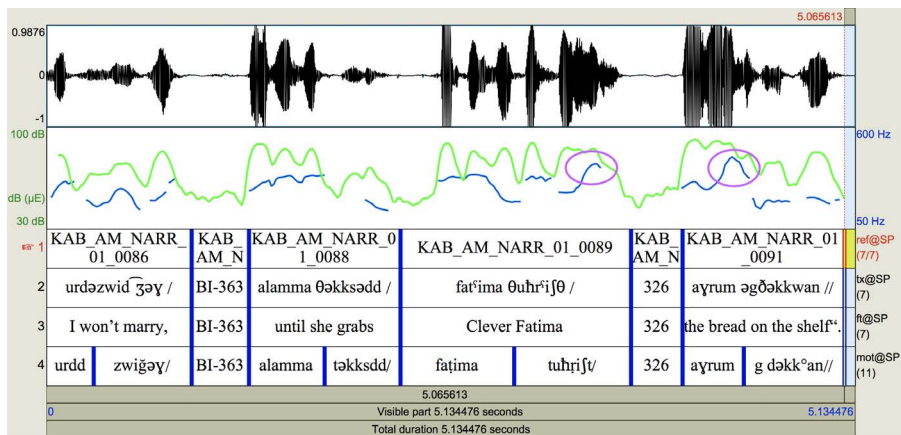
Interestingly, in example (28), as elsewhere in my data for similar examples, the continuative tone is correlated to a focal prominence on the last word of the unit, here the adverb (93 dB), and a highlighting of the direct object (with high values ranging from 89 to 91 dB): in a display of suspense and disclosure, the storyteller plays on the listener’s expectations in her rendition of the young hero’s extravagant demand to his brothers: ‘I won’t give you the partridges I hunted, for you to show our father that you are good hunters, until you each give me... one of your fingers!’

Example (1), reproduced here as (29) can also be analyzed in those terms: the prosodic group of the verb is first separated from the postverbal nominal subject, and then, after a silent pause, the nominal direct object appears, immediately followed by a locative complement.


- (29) *urdəzwid̪zəy / BI_363 / alamma θəkksədd / fatˈima θuhrˈiʃθ / 326 / aɣrum əgðəkkwan //*
 ur=dd zwiǧ-ɣ / BI-363/ alamma
 NEG=PROX marry\NEGPfV-SBJ.1SG / BI_363 / until
 PTCL=PTCL V23-PRO / BI_363 / CONJ
 t-əkks=dd / faɣima tuhrɨʃt / 326 /
 SBJ.3SG.F-take_away\PFV=PROX / Faɣima clever / 326 /
 PRO-V23=PTCL / NP ADJ / 326 /
 aɣrum g udəkk˚an //
 bread\ABS.SG.M LOC shelf\ANN.SG.M //
 N.OV PREP N.OV //
 I won't marry / ... / until she grabs / clever Fatima / ... / the bread on the shelf //
 (I won't marry until Clever Fatima grabs the bread on the shelf)
 (KAB_AM_NARR_01_0086-91)

There is a rising tone on *tuhrɨʃt* (400 Hz), with a high intensity value (85 dB), then a silent pause which adds to the highlighting effect, and then again a high F0 value on *aɣrum*(438 Hz) and high intensity as well (91 dB), as shown in Figure 3 below.

FIGURE 3 – Praat acoustic analysis of example 29 (F0 curve and Intensity)

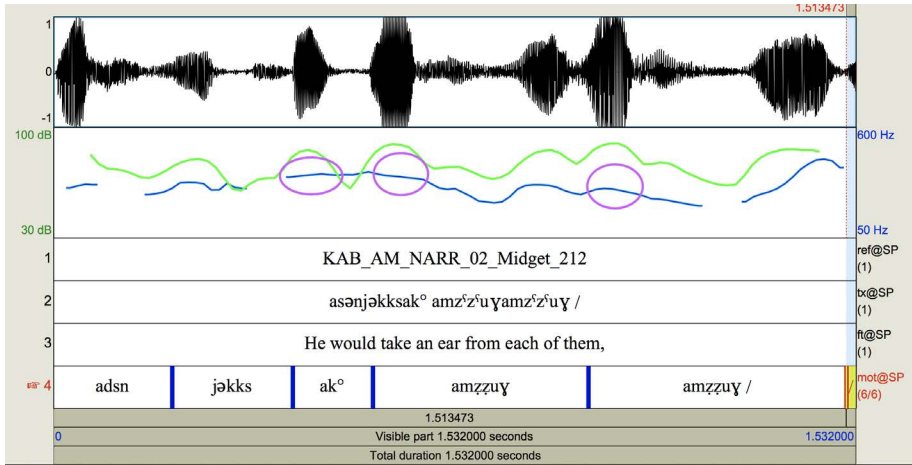


On the contrary, when the direct object is within the prosodic group of the verb, the contour is smoother, in a rising-falling curve, as in figure 3 below, corresponding to example 30:

- (30) *asənjəkksak^o amz^zz^uyamz^zz^uy /*
- | | | |
|---------------|-------------------------|-----------------|
| ad=asn | j-əkks | ak ^o |
| POT=DAT.3PL.M | SBJ.3SG.M-take_away\AOR | all |
| PTCL=PRO | PRO-V23 | ADV |
| amzzuy | amzzuy / | |
| ear\ABS.SG.M | ear\ABS.SG.M / | |
| N.OV | N.OV / | |
- He would take an ear from each of them,
KAB_AM_NARR_02_Midget_212
- 

The values for the adverb and the direct object are respectively 363 Hz/85 dB for *ak^o*, and 356 Hz/89 dB for the first *amzzuy*, and 292 Hz/89 dB for the second one.

FIGURE 4 – Praat acoustic analysis of example 30 (F0 curve and Intensity)

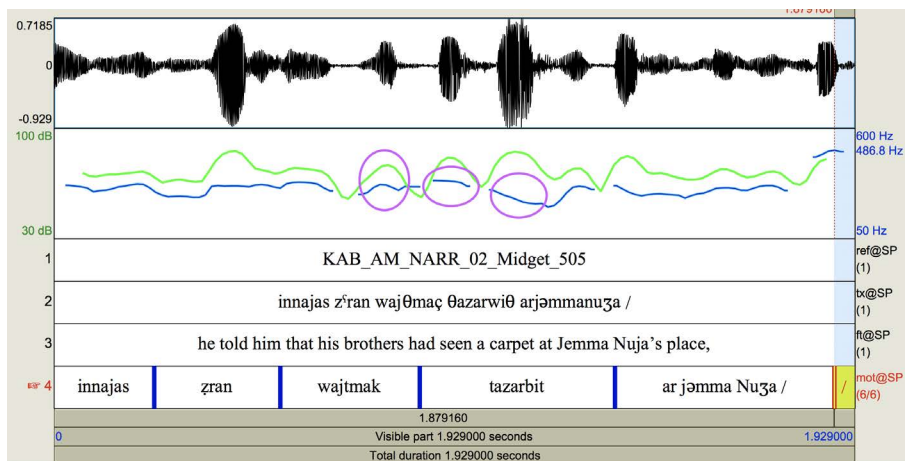


The preceding example is a special distributive construction of the direct object, chosen for its parallelism with (28), and therefore might show slightly atypical values, but the following one is quite standard:

- (31) *innajas zʳran wajθmaç θazʳarʳβiθ arjəmmanuza /*
 i-nna=as zɹa-n wajtma-k
 SBJ.3SG.M-say\PFV=DAT.3SG see\PFV-SBJ.3PL.M brother\ANN.PL.M-
 KIN.2.SG.M
 PRO-V13%=PRO V13%-AFFX N.KIN.OV-PRO
- tazərbit** ar jəmma Nuza /
carpet\ABS.SG.F to mother\ANN.SG.F Nuza /
 N.OV PREP N.KIN.COV N.P /
- 'he told him that his brothers had seen a carpet at Jemma Nuja's place'*
 (KAB_AM_NARR_02_505)

As shown in Figure 4 below, the F0 contour is smooth and slowly rising then falling, while Intensity rises slightly on the nominal subject (75 dB), then more markedly on the direct object, which is in focus (85 dB).

FIGURE 5 – Praat acoustic analysis of example (31) (F0 curve and Intensity)



Contrary to example (29), there is no peak on the element preceding the direct object, here the postverbal subject *wajtmak* (332 Hz). The value for the direct object *tazərbit* is 338 Hz on the first syllable and 239 Hz on the second one. In cases when the direct object is separated from the verb or the postverbal subject or adverb, the prosodic contour anticipates continuation, through high F0 values just before the stranded

direct object. They also show typical rhetorical features: focus is marked by high intensity values, both on the preceding element, and on the direct object itself, with sometimes anticipatory pauses as well, which increase suspense and rhetorical effect.

6. Conclusion and discussion

I have shown that the Direct Object role is marked by a dedicated construction involving a noun in the absolute, different from the semantic role of referential undergoer, which is coded by the use of an absolutive bound pronoun cliticized on a verb or verbal head.

I have given evidence for the crucial role of prosody in the formal definition of the construction, which involves prosodic boundaries: the Direct Object is a noun in the absolute state, immediately following the verb, or following <the verb followed by a noun in the annexed state> or following <the verb followed by an adverb> or following <the verb followed by a postverbal negator>, within the prosodic group of the verb.

I have shown that apparent counterexamples to that definition, namely occurrences where the noun in the absolute is detached from the prosodic group of the verb and appears in a separate Intonation Unit, in fact contain traces of a tight relationship between Verb and Direct Object: those are either disfluencies or stylistic devices, and in both cases, they contain evidence of integration between the prosodic group of the verb and the intonation unit containing the noun in the absolute: in the case of disfluencies, hesitation markers or false starts are lengthened and the sequence is immediately resumed. In the case of stylistic devices, such as anticipatory/delayed focus, a strong rising tone with continuative value informs the addressee that the prosodic group of the verb is not complete, and that the following sequence is highlighted.

In terms of method and background assumptions, the present study shows how important it is to uncover language-internal categories through the empirical study of spontaneous data, transcribed, segmented and annotated with as few aprioristic assumptions as possible. Without a notation of prosodic boundaries based on acoustic and perceptual cues rather than on syntactic or pragmatic or semantic assumptions, and without precise transcription of hesitations, false starts and pauses, it would not have been possible to conduct the investigation presented in this paper.

Moreover, the approach adopted in this study is also theoretically different from most treatments of the role of prosody in relation to grammar. I do not discard phenomena (disfluencies and stylistic devices) that are usually ascribed to ‘other levels’ of language analysis, only retaining the prosodic boundaries that are congruent with phrasal, clausal or sentential syntactic boundaries. I consider that prosodic cues are to be treated as elements of the fabric of language, just like morphological marks, linear ordering, and other formal coding means are. I do not view prosody as a separate module, and intonation units as a projection of other structural levels of grammar, or as a pragmatic unit with a single functional value (speech-act or other). My findings plead for an integrated view of prosody, closely interacting with syntax, semantics, phonology, information structure, and all levels of human communication and cognition, in a way that is best represented as a complex weaving of various threads, rather than a piling up of neatly stacked and hierarchically organized layers. I suggest that what linguists have first separated into different domains in order to be able to address problems in a structured, progressive and modular way, should not be reified into a representation of what language actually is. The various forms that we are able to isolate as elements contributing to the construction of meaning, are in fact part of a complex whole whose interrelations are still to be fully understood.

Acknowledgements

I am grateful to the two anonymous reviewers for their appreciation and comments, and hope, thanks to their suggestions, to have improved the explicitation of the approach and methodology so that those can be tested on other languages.

References

CHAFE, W. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: The University of Chicago Press, 1994.

CRESTI, E.; MONEGLIA, M. *C-Oral-Rom*. Integrated Reference Corpora for Spoken Romance Languages. Amsterdam/Philadelphia: John Benjamins, 2005.

CRUTTENDEN, A. *Intonation*. Second edition. Cambridge: Cambridge University Press, 1997.

IZRE'EL, S.; METTOUCHI, A. Representation of speech in CorpAfroAs: Transcriptional strategies and prosodic units. In: METTOUCHI, Amina; VANHOVE, Martine; CAUBET, Dominique (Coord.). *Corpus-based Studies of Lesser-described Languages: The CorpAfroAs corpus of spoken Afroasiatic languages*. Amsterdam; Philadelphia: John Benjamins, 2015. p. 13-41. Doi: <http://dx.doi.org/10.1075/scl.68>

MATTHEWS, P. *The Concise Oxford Dictionary of Linguistics*. Oxford: Oxford University Press, 2007.

METTOUCHI, A. *Segmenting spoken corpora in lesser-described languages: new perspectives for the structural analysis of speech*. In: ANNUAL MEETING OF THE SOCIETAS LINGUISTICA EUROPAEA, 46th., Split (Croatia), 18-21 Sept. 2013. Plenary talk.

METTOUCHI, A.; FRAJZYNGIER, Z. A previously unrecognized typological category: The state distinction in Kabyle (Berber). *Linguistic Typology*, De Gruyter, v. 17, n. 1, p. 1-30, 2013. Doi: 10.1515/lity-2013-0001

METTOUCHI, A. Foundations for a typology of the annexed/absolute state systems in Berber. *Language Typology and Language Universals (STUF)*, Berlin; Boston, v. 67, n. 1, p. 47-61, 2014. Special issue: TAINE-CHEIKH, Catherine; LUX, Cécile (Ed.). *Berber in Typological Perspective*. DOI: <https://doi.org/10.1515/stuf-2014-0005>

METTOUCHI, A. Aspect-Mood and discourse in Kabyle (Berber) spoken narratives. In: PAYNE, Doris L.; SHIRTZ, Shahar (Ed.). *Beyond Aspect: the expression of discourse functions in African languages*. Amsterdam; Philadelphia: John Benjamins, 2015. p. 119-143. Doi: <https://doi.org/10.1075/tsl.109>

METTOUCHI, A. Predication in Kabyle (Berber). In: METTOUCHI, A.; FRAJZYNGIER, Z.; CHANARD, C. (Ed.). *Corpus-based cross-linguistic studies on Predication (CorTypo)*. 2017. Available at: <http://cortypo.huma-num.fr/>.

METTOUCHI, A. The interaction of state, prosody and linear order in Kabyle (Berber): Grammatical relations and information structure. In: TOSCO, Mauro (Ed.). *AfroAsiatic: Data and Perspectives*. Amsterdam: John Benjamins, 2018. p. 261-285. (Current Issues in Linguistic Theory 339). [Paper first presented at the 14th Italian Meeting of AfroAsiatic Linguistics in Turin, on 17 June 2011]). Doi: <https://doi.org/10.1075/cilt.339>

PIERREHUMBERT, J.; HIRSCHBERG, J. The Meaning of Intonational Contours in the Interpretation of Discourse. In: COHEN, P.; MORGAN, J.; POLLACK, M. E. (Coord.). *Intentions in Communications*. Cambridge, Mass.: MIT Press, 1990. p. 271-311.

ROSS, B. *Prosody and Grammar in Dalabon and Kayardild*. 2011. Thesis (PhD) - University of Melbourne, Australia, 2011.

SCHEER, T. Chunk definition in phonology: prosodic constituency vs. phase structure. In: BLOCH-TROJNAR, M.; BLOCH-ROZMEJ, A. (Coord.). *Modules and Interfaces*. Lublin: Wydawnictwo KUL, 2011. p. 221-253.

SELKIRK, E. On Clause and Intonational Phrase in Japanese: The Syntactic Grounding of Prosodic Constituent Structure. *Gengo Kenkyu* (Journal of the Linguistics Society of Japan), n. 136, p. 35-74, 2009.

TAO, H. *Units in Mandarin Conversation: Prosody, Discourse, and Grammar*. Amsterdam; Philadelphia: John Benjamins, 1996.

VOGEL, I. Phonological Words. In: BROWN, K. (Coord.). *Encyclopedia of Language and Linguistics*. 2. ed. Oxford: Elsevier, 2006. p. 531-534.



Prosody and Processing: Comprehension and Production of Topic-Comment and Subject-Predicate Structures in Brazilian Portuguese

Prosódia e processamento: compreensão e produção de estruturas de tópico e de sujeito no português brasileiro

Andressa Christine Oliveira da Silva

Universidade Federal de Juiz de Fora, Juiz de Fora, Minas Gerais / Brasil
andressa.silva@letras.ufjf.br

Aline Alves Fonseca

Universidade Federal de Juiz de Fora, Juiz de Fora, Minas Gerais / Brasil
aline.fonseca@letras.ufjf.br

Abstract: This paper explores the influence of prosody in the processes of comprehension and production of sentences in Brazilian Portuguese with topic-comment syntactic structure and sentences with subject-predicate syntactic structure, in active or passive voice. Three experimental activities were carried out, one production task and two comprehension tasks. Experiment 1 consisted of a perception task with the ABX technique, and it aimed to test if hearers recognize prosodic differences between topicalized Determinant Phrases (DPs) and DPs in subject position. Experiment 2 consisted of a sentence elicitation task with Cross-modal naming technique and it aimed to investigate whether Portuguese native speakers produce a subject-predicate structure or a topic-comment structure in contexts that favor the occurrence of these syntactic structures in speech. Experiment 3 consisted of a comprehension task with Self-paced listening and reading technique and it aimed to investigate whether prosodic characteristics of a DP, in topic or subject position, can guide hearers during the processing in order to distinguish between these two syntactic categories. From the comprehension/perception perspective, the results of the experiments 1 and 3 indicated that speakers recognize the prosodic differences between the topicalized DPs and the subject DPs, and use such characteristics during linguistic processing. From

the production perspective, the results of experiment 2 revealed that speakers are able to produce sentences consistent with topic-comment and subject-predicate syntactic structures when the context favors the occurrence of one of them. Nevertheless, the results also reveal a preference for the subject-predicate structure over the topic-comment structure in BP.

Keywords: prosody-syntax; topic-comment; subject-predicate.

Resumo: Este trabalho investiga a influência da prosódia nos processos de compreensão e produção de sentenças com elementos topicalizados, do tipo tópico-comentário, e sentenças com a estrutura de sujeito-predicado, na voz ativa ou passiva, do Português Brasileiro. Aplicaram-se três atividades experimentais, uma tarefa de produção e duas de compreensão. O Experimento 1 consistiu em um teste de percepção com a técnica ABX, cujo objetivo foi testar se ouvintes reconhecem as diferenças prosódicas entre *Determinant Phrases* (DPs) topicalizados e DPs em posição de sujeito não topicalizado. O Experimento 2 consistiu em um teste de elicitación de frases com imagens do tipo *Cross-modal naming*, cujo objetivo foi investigar se em contextos que favorecem a ocorrência de estruturas de sujeito ou de estruturas topicalizadas, os falantes produzem frases consistentes com tais estruturas sintáticas. O Experimento 3 consistiu em uma tarefa de compreensão, com a técnica *Self-paced listening and reading*, cujo objetivo foi investigar se as características prosódicas de um DP, em posição de tópico ou de sujeito, conseguem guiar o processamento linguístico dos ouvintes na distinção entre essas duas categorias sintáticas. Na compreensão/percepção, os resultados dos experimentos indicaram que os falantes reconhecem as diferenças prosódicas entre os DPs topicalizados e os DPs em posição de sujeito, e utilizam tais características durante o processamento linguístico. Na produção, os resultados revelaram que os falantes produzem frases consistentes com estruturas sintáticas de tópico e de sujeito quando o contexto favorece o aparecimento delas, entretanto, apontam para uma preferência da estrutura de sujeito como *default* no PB.

Palavras-chave: prosódia-sintaxe; tópico-comentário; sujeito-predicado.

Submitted on December 10th, 2017

Accepted on February 27th, 2018

1 Introduction

This work presents the research findings of a master's dissertation (SILVA, 2017) that explored sentences in Brazilian Portuguese (hereafter BP) formed by the topic-comment syntactic structure, which presents the

internal argument of the verb at left-edge of the clause, and sentences formed by the subject-predicate syntactic structure in active or passive voice, as in the examples shown below:

(1) **Topic-Comment**

[*A mochila vermelha*]_{Topic} [*Ana comprou no shopping*]_{Comment}
 [The red backpack]_{Topic} [Ana bought (it) in a shopping mall]_{Comment}

[*A menina*]_{Topic} [*a tia levou no shopping*]_{Comment}
 [The girl]_{Topic} [the aunt took (her) to the shopping mall]_{Comment}

(2) **Subject-Predicate: Passive Voice**

[*A mochila vermelha*]_{Subject} [*foi comprada no shopping*]_{Predicate}
 [The red backpack]_{Subject} [was bought in a shopping mall]_{Predicate}

(3) **Subject-Predicate: Active Voice**

[*A menina*]_{Subject} [*esperou o pai na portaria*]_{Predicate}
 [The girl]_{Subject} [waited for her dad at the entrance]_{Predicate}

One of the reasons to choose the topics as object of study is the fact that these syntactic structures present particular prosodic characteristics (MORAES; ORSINI, 2003), which distinguish them from the subject-predicate structure. The position of the topic is at the beginning of the sentence, it announces what the theme of the statement is. The comment brings what is said about the topicalized element. When the topic is moved to the beginning of the sentence it leaves the root sentence¹ and forms a single intonational phrase, or IP (see Prosodic Hierarchy of NESPOR; VOGEL, 2007). A topic-comment sentence tends to be formed by two IPs and between the topic and the comment there is usually the occurrence of a pause. The subject-predicate structure, on the other hand, tends to form only one IP, which does not favor the occurrence of pauses between the elements.

The second reason to explore the topics is the fact that there are few studies in BP that investigate these constructions through an

¹ The root sentence is understood as a single [NP VP]-structure without extrapositions or interruptions (GUSSENHOVEN; JACOBS, 2011, p. 252).

experimental perspective (KENEDY, 2011, 2014; SILVA, 2015), in order to identify how speakers process these structures. Most of the researches in BP study these constructions by using spoken corpora and they explore mainly their discursive and syntactic characteristics over their prosodic aspects.

The third reason is due to the fact that there are few studies in the prosody-syntax interface that investigate whether prosodic information can also guide the processing of syntactic structures without interpretative ambiguities. In Psycholinguistics, many studies in the prosody-syntax interface have investigated the role of prosody in the disambiguation of syntactic structures (CARLSON *et al.*, 2001; CLIFTON JR. *et al.*, 2002; FRAZIER *et al.*, 2003; among others).

Finally, there is also an uncertainty in the linguistic literature about the status of BP in the typology of languages proposed by Li and Thompson (1976). There are linguists who claim that spoken BP is both subject-prominent and topic-prominent (PONTES, 1987; ORSINI, 2003; among others) and there are other linguists who claim that spoken BP is a subject-prominent language (KENEDY, 2011, 2014; among others).

Considering the reasons presented previously, the main goal of this research is to investigate the role of prosody in the processes of comprehension and production of topic-comment and subject-predicate structures through experimental evidences. As specific objectives, we intend to: (i) analyze the prosodic characteristics present in topic-comment structures and those present in subject-predicate structures; (ii) verify if native BP speakers recognize prosodic differences between a DP in the position of topic and a DP in the position of non-topicalized subject; (iii) identify whether there is a preference in spoken language for one of the two structures; (iv) investigate whether the prosodic characteristics of a topic DP or a subject DP are sufficient and informative to guide the linguistic processing towards the distinction between these two syntactic categories; (v) verify if hearers recognize when there is a mismatch between the prosodic structure and the syntactic structure in topic-comment sentences and in subject-predicate sentences. In order to fulfill these objectives, three experimental tasks were designed: a perception task with ABX technique, a production task with Cross-modal naming technique and a comprehension task with Self-paced listening and reading techniques.

2 Theoretical background

In a classic study in the descriptive literature about topics, Li and Thompson (1976) claimed that every language has the topic-comment construction; however, languages differ in relation to the strategies used to construct sentences. The researchers analyzed spoken corpora of several languages taking into account the strategies in the construction of sentences according to the prominence of the notions of subject and topic. They found out four basic types of languages:

- (i) Languages that are subject-prominent.
- (ii) Languages that are topic-prominent.
- (iii) Languages that are both subject-prominent and topic-prominent.
- (iv) Languages that are neither subject-prominent nor topic-prominent.

(Adapted from LI; THOMPSON, 1976, p. 459)

In type (i) languages, English for instance, the grammatical relation subject-predicate plays a major role. In type (ii) languages, such as Chinese, the basic structure of sentences favors the grammatical relation of topic-comment. In type (iii) languages, Japanese for instance, there are two sentence construction strategies that are equally important, both topic-comment and subject-predicate. In type (iv) languages, such as Tagalog, the notions of topic and subject have merged to such an extent that it is no longer possible to distinguish them in any type of sentence.

The authors outlined seven differences between subjects and topics in terms of properties they do not share. They are summarized below:

- (a) **Definite:** The topic must be definite while the subject need not be definite, it might be indefinite.
- (b) **Selectional relations:** The topic need not have a selectional relation with any verb in a sentence, that is, it need not be an argument of a predicative constituent. The subject, on the other hand, is always selectionally related to some predicate in the sentence.

- (c) **Verb determines “Subject” but not “Topic”:** A correlate of the fact that a subject is selectionally related to the verb is the fact that, with certain qualifications, it is possible to predict what the subject of any given verb will be. The topic, on the other hand, is not determined by the verb; topic selection is independent of the verb. Discourse may play a role in the selection of the topic.
- (d) **Functional role:** The functional role of the topic is constant across sentences. It specifies the domain within which the predication holds. Thus, the topic is the “center of attention”; it announces the theme of the discourse. This is why the topic must be definite. Looking at the functional role of the subject, on the other hand, reveals two facts. First, some NPs do not play any semantic role in the sentence at all; that is, in many subject-prominent languages, sentences may occur with “empty” subjects. Second, in case the subject NP is not empty, the functional role of the subject can be defined within the confines of a sentence as opposed to a discourse.
- (e) **Verb-agreement:** The verb in many languages shows obligatory agreement with the subject of a sentence. Topic-agreement, however, is very rare. Topics are much more independent of their comments than are subjects of their verbs.
- (f) **Sentence initial position:** Although the surface coding of the topic may involve sentence position as well as morphological markers, it is worth noting that the surface coding of the topic in all the languages involve the sentence-initial position. Subject, on the other hand, is not confined to the sentence-initial position. The reason that the topic but not the subject must be in sentence-initial position may be understood in terms of discourse strategies.
- (g) **Grammatical processes:** The subject but not the topic plays a prominent role in such processes as reflexivization, passivization, Equi-NP deletion, verb serialization, and imperativization. These processes are concerned with the internal syntactic structure of sentences. Since the topic is syntactically independent in the sentence, it does not play a role in the statement of these processes.

(Adapted from LI; THOMPSON, 1976, p. 461-466)

The researchers emphasize that these seven criteria are not intended to constitute a definition of the notion of subject or topic, but are designed to serve as guidelines for distinguishing topics from subjects. Overall, these criteria point out that the topic is a discourse notion, whereas the subject is more related to a sentence-internal notion. The topic can be understood best in terms of discourse and extra-sentential considerations, while the subject can be best understood in terms of functions within the sentence structure.

Besides the characteristics that differentiate the topic from the subject, Li and Thompson also present some characteristics that are typical of topic-prominent languages:

- (a) **Surface coding:** In topic languages, there is a surface coding for the topic, such as a morphological marker, for instance.
- (b) **The passive construction:** Among topic languages, passivization either does not occur at all, or appears as a marginal construction, rarely used in speech, or carries a special meaning.
- (c) **“Dummy” subjects:** “Dummy” or “empty” subjects, such as the English it and there, the German es, the French il and ce, are not found in topic languages.
- (d) **“Double subject”:** Topic languages are famous for their pervasive so-called “double subject” construction. Such sentences are the clearest cases of topic-comment structures.
- (e) **Controlling co-reference:** The topic typically controls co-referential constituent deletion.
- (f) **V-final languages:** Topic languages tend to be verb-final languages.
- (g) **Constraints on topic constituent:** In topic-comment languages, there are no constraints on what may be the topic.
- (h) **Basicness of topic-comment languages:** In topic languages, the topic-comment sentence can be considered to be part of the repertoire of basic sentence types.

(Adapted from LI; THOMPSON, 1976, p. 466-471)

Li and Thompson state that in the search for linguistic universals the typology of languages proposed by them can really serve as a description of strategies for achieving this goal.

Turning to BP, the pioneering work of Pontes (1987) seeks to identify which type of language spoken BP is in Li and Thompson's typology. According to her, BP has always been considered a subject-prominent language in linguistic literature, however, she emphasizes that studies about spoken BP were scarce. When she observed the spontaneous and colloquial language in ordinary usage, it was verified that topic-comment structures are widely recurrent in spoken language. She also points out that these constructions are of different types. Pontes claims that NPs with different functional roles can constitute a topic in BP: indirect object, direct object, adjuncts, complements, subjects. According to her, the most frequent type of topic construction in spoken BP is "Books, they are on the table" (1987, p. 12), which can occur with or without a pause after the topic NP.

Pontes explored spoken data in order to classify BP in Li and Thompson's typology. She analyses her database according to the seven criteria to differentiate topic from subject and the typical characteristics of languages with prominence of topics. Examples for all the seven criteria were found. Regarding the typical characteristics of topic languages, the researcher found out that BP, with the exception of the surface coding feature, presents all the other characteristics of topic-prominent languages. One aspect noticed by her in the database was the occurrence of a co-referential pronoun to refer to the topic, also known as pronoun copy. The presence of the pronoun copy is much greater when the topic is identical to the subject of the comment sentence than when it refers to other elements of the sentence. She points out that the greater incidence may be due to the difficulty in distinguishing whether the subject, when in sentence-initial position, is also a topic or only a subject. However, she affirms that this is not the only function of the pronoun copy. In other cases, the presence of this pronoun can be accounted by the distance between the topic/subject and the verb to which it is attached. Due to the necessity for making clearer what the referent is, the speaker would use this pronoun copy. On the other hand, in the examples of sentences in which the topic refers to other constituents, the occurrence of this pronoun is less frequent. It appears in cases of difficulty to identify the referent, to give emphasis or to contrast. Pontes emphasizes that in colloquial BP

the verbal inflection forms are diminishing and, consequently, it becomes more difficult to identify the referent, since a certain verbal inflection can refer to different people in discourse. In these cases, the pronoun copy would help to identify the referent.

There was another aspect explored by Pontes (1987), which was related to the nature of topic constructions. She claims that Ross (1967) established a distinction between the topic constructions that are generated by Left Dislocation (LD) and those that are generated by Topicalization (TOP) in American English. In the former there is the occurrence of a pronoun copy, for instance “The man my father works with, he’s going to tell the police that...” (*O homem que trabalha com meu pai, ele vai dizer à polícia que...*). In the later the pronoun-copy does not appear, such as in “Beans I don’t like” (*Feijões eu não gosto*). In BP, however, Pontes states that it is difficult to apply this distinction since it is possible to omit the pronoun. Overall, pronoun omission in BP is always possible if there is no impairment of meaning. Therefore, the fact that the pronoun is optional makes it difficult to identify if it is a TOP construction or an LD construction with elided pronoun. The author analyzes several examples in her database in order to reach a possible distinction between LD and TOP, but she does not find a conclusion that there would be a difference between the two constructions in BP. She points out that it is tempting to make distinctions between the two constructions, for clear cases, in the following way:

TABLE 1 – Topicalization and Left Dislocation features (PONTES, 1987, p. 82)

Topicalization features	Left Dislocation features
No pause	Pause
No pronoun copy	Pronoun copy
Contrastive	Non-contrastive
Definite NPs or Indefinite NPs	Definite NPs

Pontes argues, however, that due to the cloudiness of the phenomenon, it would be premature to decide on the distinction between the two types of construction until the conditions of pronominalization as well as elision of pronouns in BP are more explored. A broader study

about topic constructions in speech could also be helpful to clarify the phenomenon.

Overall, all the aspects investigated in her dataset suggest to Pontes that BP should be considered at least a type (iii) language in Li and Thompson's typology, in which both subject-predicate and topic-comment constructions are prominent.

In relation to the current research, the type of topic-comment construction adopted varies according to the experimental aims, that is, they could present features of the two types of topic constructions defined by Pontes (1987). For experiments 1 and 3, stimuli were designed with the features of both TOP and LD; there was the topicalization of the internal argument of the verb of the comment sentence, without the occurrence of a pronoun copy, but with the occurrence of a pause between the topic and the comment. In experiment 2, stimuli could match the features of both TOP and LD, depending on participants' production choices.

With regard to prosodic aspects, the research conducted by Callou *et al.* (1993) was one of the first works in BP to explore the topics in the prosody-syntax interface. In that work, the authors observed syntactic and prosodic features present in TOP, LD and subject-predicate constructions in spoken data. The analyses revealed that the most frequent prosodic pattern for TOP is rising intonation, while for LD a balanced distribution of the patterns was observed. In proportional terms, the falling intonation was more frequent for LD than for TOP. In relation to pause, TOP and LD constructions present similar distribution of long pauses and average pauses. Regarding the micropauses, TOP presented a greater occurrence of them over LD. Although there was no marked polarization in all observed cases, TOP and LD differed in relation to the direction of the melody curve. However, the distinction between the two constructions was less marked when the intonational curve was treated separately from the pause. The authors found out that prosody was only distinctive when the opposition was made between topic-comment and subject-predicate structures.

In summary, they concluded that prosody could not clearly distinguish TOP from LD, since the diversity of patterns found for TOP – intonational curve and pause – was also found for LD. They also affirm that the lack of a pattern that only occurs with topic-comment leads them to believe that focus marking in this type of construction is little used. Therefore, the distinction between TOP and LD would have

complementary distribution, based on a grammatical conditioning, and not on a prosodic one.

Orsini (2003) also conducted a research about topics in the prosody-syntax interface. Her work explored two main aspects, the syntactic and discursive features of topic structures and their prosodic features. She found out four types of topic construction strategies in the spoken corpora, however, our work is going to focus only on TOP and LD constructions. The database revealed that there were more TOP constructions than LD constructions. Regarding prosodic analyses, the author found out three distinct prosodic patterns. The subject-predicate sentences presented mostly the prosodic pattern H* L+H* H%, which was also observed in most topicalization constructions, regardless of the syntactic value of the topic. The LD constructions presented mostly the intonational pattern H* L+H* L% followed by a pause. When the topic presented contrastive value, in both TOP and LD constructions the prosodic pattern L* H*+H H% was observed. No pause was found between the topic and the comment. The author points out that these three prosodic patterns are not exclusive for topic construction strategies because they also occurred with the four types of topic construction strategies. Therefore, there are no exclusive intonational patterns and there are no topic construction strategies that reveal categorical intonational patterns.

In summary, Orsini concluded that intonational patterns differentiate TOP structures from LD structures, however, she did not detect any significant prosodic features that differentiate subject-predicate sentences from topicalization sentences. She also points out that the results are in line with Callou *et al.* (1993), in the sense that there is no exclusive intonational pattern for each topic construction strategy. On the other hand, this result points to the existence of three distinct and systematic prosodic patterns, which leads her to defend that BP reveals two independent modules – one syntactic and one prosodic – that interact with one another. Orsini, in the same line as Pontes (1987), also claims that BP should be considered a type (iii) language in Li and Thompson's typology.

With regard to sentence processing, three researches are outlined here, the works of Kenedy (2011, 2014) and the work of Silva (2015). Kenedy (2011, 2014) points out that it is relevant to approach the cognitive processing of topic-comment structures as opposed to subject-

predicate structures through an experimental perspective, since there are few studies in BP that explore these constructions in experimental tasks. He claims that most researches have investigated the BP status in Li and Thompson's typology (1976) based on spoken data. The author believes that this type of methodology is limited, since the results could be strongly biased by the subjects' sociocultural profile and/or by the textual genre under investigation. The experimental methodology, on the other hand, could indicate interesting results about the typological status of BP, since the tests are carried out in controlled laboratory situations and the results are submitted to reliable statistical tests. Therefore, Kenedy (2011, 2014) conducted three experimental activities in total: a self-paced reading task, a self-paced listening and a speeded judgment task, to compare the processing of topic-comment structures against subject-predicate structures.

In the self-paced reading task, the author explored sentences that presented as first segments initial DPs, which could be interpreted initially as a topic or as a subject. Only when participants had read the second segment, which presented a VP, they could attribute to the sentence a mental representation of topic structure or subject structure. It is worth mentioning that this type of topic structure explored by Kenedy (2011, 2014) is classified as topic-subject by some authors, such as Orsini (2003) and Callou *et al.* (1993). In this type of construction, the topic is reanalyzed as a subject, and the verb agreement is established, which contributes to the maintenance of the SVO canonical order of BP. The results of this experiment indicated that participants spent more time reading the critical segment (the VP) of the sentences in the topic condition compared to the sentences in the subject condition. Therefore, the topic structure was cognitively more costly to process than the subject structure. For Kenedy, this result contradicts the hypothesis that BP is a language with prominence of topics.

A self-paced listening experiment was designed to verify if the absence of prosody influenced the results of the reading task. The same stimuli explored in the previous task were used in this task. For topic stimuli, there were two types of condition, one with prosody typical of a topic structure and another with prosody typical of a subject structure. The results indicated that participants had spent more time listening to the critical fragment of the topic condition with no specific melodic contour than to the subject condition. On the other hand, when the topic

condition presented typical melodic contour of topicalization, reponse times decreased considerably if compared to the other topic condition, and they are also similar to the average reponse times of the subject condition. Kenedy points out that the results of this experiment do not invalidate the hypothesis that BP is a language with prominence of topics, because when the topic structures had specific prosodic cues, they presented reaction times similar to those observed for the subject condition.

The speeded judgment task was designed to explore the phenomenon of anaphoric co-reference. The aim was to verify what the preference of Brazilian speakers is when they have to assign a lexical pronoun or an empty category to a nominal constituent that occupies either the topic position or the subject position in a sentence. In this task, subjects had to read a set of sentences and after reading each sentence, they had to rank the sentence read as acceptable or unacceptable. The results of this experiment showed that participants prefer DPs in topic position to take up a null anaphor, while DPs in the subject position should be taken up by a full pronominal anaphor. Regarding reaction times, the results indicated that the topic conditions demanded more time on the judgment than the subject conditions. The author argues that together these results also refute the hypothesis that BP is a language with prominence of topics.

Overall, the results of the three tasks revealed that it is cognitively more costly for speakers to process the topic-comment structures over the subject-predicate structures. For the researcher, this result counters the claim that BP is a subject-prominent and topic-prominent language (PONTES, 1987; ORSINI, 2003).

Silva (2015) explored in the prosody-syntax interface whether prosody is able to guide the syntactic processing of topic-comment and subject-predicate structures. She conducted two production tasks and three comprehension tasks.

The first production task consisted of naive subjects reading sentences aloud for recording. First, participants had to read a sentence without having read it beforehand. Then they should read that same sentence again two more times. The experimenter analyzed just the first and third readings. The number of times each sentence was read with either the prosody of topic or the prosody of subject was counted. The results indicated that in the first reading, in which participants did not know the meaning of the sentences, they preferred mainly the prosody

of subject. On the other hand, in the last reading, in which participants already knew the meaning of the sentences, most sentences with topic structure were read with the prosody of topic. Therefore, these results point out that the prosody of the subject structure seems to be the default in BP, whenever there is no previous knowledge of the sentence. Regarding the intonational characteristics, in the first reading of the sentence with topic structure, a prosody of subject was found, with the L+H* L% pattern. In the third reading, a topic prosody was found, with the H+L* H% pattern in the topicalized constituent. The second production task also consisted of reading sentences aloud for recording; however, it was done by a participant who knew the aims of the research. This task was conducted in order to verify if there are any prosodic differences between topic-comment sentences and subject-predicate sentences. In the topic structure, there was an IP boundary signaled by a long pause after the topicalized constituent, lengthening of the stressed syllable of the topicalized constituent and a descending melodic contour at the end of the sentence. In the subject structure, there was a boundary between the name and the verb, signaled by a shorter pause, there was lengthening of the stressed syllable of the name in the subject position, and a descending melodic contour signaling the end of the sentence. The results revealed that there are different prosodic structures depending on the type of syntactic structure.

The comprehension tasks consisted of a speeded judgment experiment and two self-paced listening experiments. They were designed in order to find out how hearers perceive the prosodic cues and how such cues can guide the sentence processing. The speeded judgment task sought to verify the naturalness of topic-comment sentences and subject-predicate sentences recorded both in the cooperating prosody version and in the baseline prosody version. In this task, after listening to each of the sentences participants had to judge them as: (a) unnatural; (b) not very natural; or (c) natural. The results showed that participants preferred the topic sentences in the cooperating prosody version than in the baseline version. The author claims that this result is due to the fact that the topic structure is more marked and more context-dependent.

In the first self-paced listening task, the topic and the subject sentences were presented in the cooperating prosody version. The aim was to evaluate if hearers would be able to perceive a mismatch between the prosodic structure and the syntactic structure, that is, the

initial constituent of the topic sentences were replaced with the initial constituent of the subject-predicate sentences and vice versa. Topic and subject conditions in which prosody and syntax matched were also presented. The results indicated that participants only identified a mismatch in topic conditions with incongruence between prosody and syntax. The author states that because the topic structure is more marked and more discourse dependent, it would also sound more natural with a more prominent prosody than with a neutral subject prosody. In the case of the subject-predicate structure, she believes that because it is the default in BP, it does not suffer as much influence from the prosodic information as the topic structure does. In the second self-paced listening experiment, Silva explored only the sentences in the baseline version, in order to investigate whether in the absence of relevant prosodic cues, hearers processed the structure preferentially as subject-predicate or as topic-comment. In this task, the two conditions had similar syntactic and prosodic structures up to the second segment. Only when they had listened to the third segment the ambiguity could be solved. The results indicated that the reaction times of the third segments were higher in the topic condition than in the subject condition. The researcher concludes that the baseline prosody led hearers to perceive the ambiguous structure preferentially as subject-predicate. When they encountered the topic structure a strangeness occurred, being necessary to reanalyze the sentence. This reanalysis manifested itself in the larger reaction times observed in the topic condition.

To summarize, the results of the production experiments revealed that there are acoustic cues that differentiate the two types of structure. In addition, they also suggested that there is a preference for subject-predicate prosody when the participant is unaware of the full meaning of the sentences. In the comprehension tasks, she found out that prosody can guide the parser in the formulation of the syntactic structure, providing cues for the construction of the syntactic structure in the course of sentence processing.

3 Experiment 1: ABX task

The ABX task consists in presenting three auditory stimuli A, B and X. Stimuli A and B differ by some quantitative difference, and stimulus X can be matched to either A or B (BOLEY; LESTER, 2009).

In this research, the experiment was designed to investigate whether speakers perceive prosodic differences between topicalized DPs and subject DPs and whether they are also able to match these DPs to sentences that present DPs with the same prosodic characteristics.

3.1 Materials

Stimuli were constructed according to a design 2x2: (i) type of syntactic structure: topic-comment syntactic structure and subject-predicate syntactic structure; and (ii) initial DP size: seven-syllable DP and ten-syllable DP. This design permitted the construction of four conditions, which were named as: Topic Condition (TC), Subject Condition (SC), Long Topic DP Condition (LTC), and Long Subject DP Condition (LSC).

The topic-comment sentences had an initial DP (with seven or ten syllables), which was the internal argument of the verb of the comment sentence. The comment sentence had a proper noun (feminine or masculine), a verb followed by another DP or a Prepositional Phrase (PP); both DP and PP could have syntactic function of indirect object or adjunct. The subject-predicate sentences, which were in passive voice, had an initial DP (with seven or ten syllables) followed by the passive structure (be + past participle) and a PP. Although the passive structure could be considered a type of topicalization in the literature (PERINI, 2010), it was used in order to keep the linguistic material similar to the linguistic material of the topic-comment sentences. Examples of the four conditions are shown below:

- (4) **Topic Condition (TC):** O álbum de retratos, Alice guardou na gaveta.
The portrait album, Alice kept in the drawer.
- (5) **Subject Condition (SC):** O álbum de retratos foi guardado na gaveta.
The portrait album was kept in the drawer.
- (6) **Long Topic DP Condition (LTC):** O álbum de retratos da festa, Alice guardou na gaveta.
The party portrait album, Alice kept in the drawer.

- (7) **Long Subject DP Condition (LSC):** O álbum de retratos da festa foi guardado na gaveta.

The party portrait album was kept in the drawer.

Sixteen sentences for each condition were constructed, sixty-four in total. The sentences were recorded by a female native speaker of BP, with training in ToBI analysis and experience recording experimental sentences. After the recording, the software Praat (BOERSMA; WEENICK, 2008) was used to isolate the initial DPs of the stimuli. There was a 100-millisecond manipulated pause after the initial DPs in Topic Conditions – TC and LTC. The initial DPs in Subject Conditions – SC and LSC – did not present pauses.

Besides the stimuli, additional twenty-eight sentences were created. Those sentences presented different types of syntactic structures. Four of them were chosen to compose the training session. They were recorded in two versions: in the first version, the sentences were read with a baseline prosody, whereas in the second version, one of the constituents of those sentences was read with focus. Subsequently, the constituent, which was read with neutral prosody in the first version and read with focus in the second version, was isolated.

3.1.1 Prosodic characteristics of stimuli

The prosodies will be described using the ToBI transcription system (PIERREHUMBERT, 1980; BECKMAN; PIERREHUMBERT, 1986) and according to Prosodic Phonology Theory (NESPOR; VOGEL, 2007).

The topicalized DPs were within a single intonational phrase (IP) and they showed a pre-nuclear accent LH on the first phonological word and a pitch accent L+H* on the last phonological word. A high boundary tone H% was also found. The comment sentences, which were within the second IP, showed a pitch accent H+L* on the last phonological word and a final low boundary tone L%. These prosodic characteristics are typical of broad-focus statements in BP (FROTA *et al.*, 2015). Regarding durational measurements, the last phonological word of topicalized DPs showed lengthening of the nuclear and the post-nuclear syllables (FONSECA, 2012). Between the topicalized DP and the comment sentence there was a 100-millisecond manipulated pause. Spectrograms of stimuli in Topic Conditions (Figures 1 and 2) are shown

below. The red circles represent the DPs that were isolated to compose stimuli A and B in the task.

FIGURE 1 – Long Topic DP Condition (LTC): pitch track for item *The party portrait album, Alice kept in the drawer*

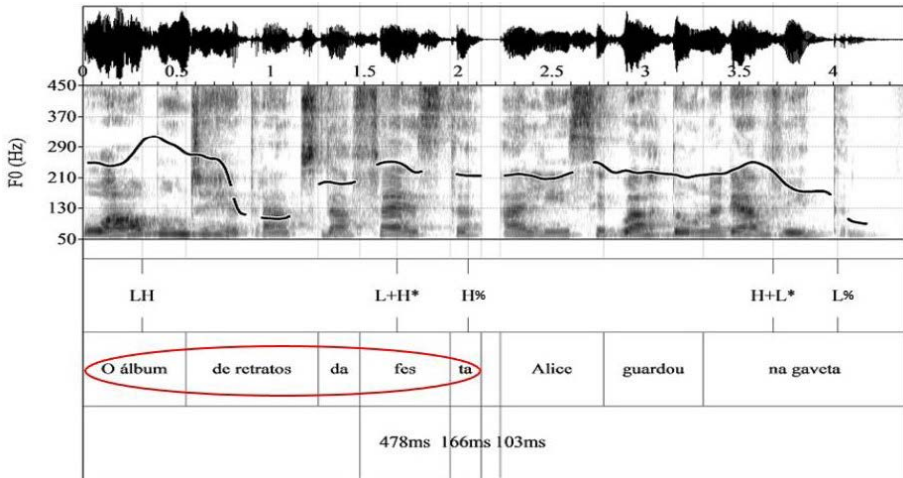
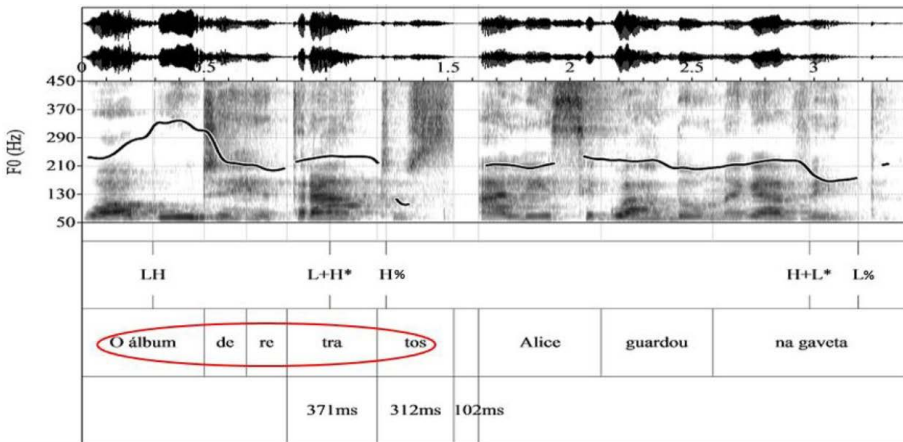


FIGURE 2 – Topic Condition (TC): pitch track for item *The portrait album, Alice kept in the drawer*



The sentences in Subject Conditions were within a single IP and showed typical characteristics of broad-focus statements. The Long Subject DP Condition (LSC) showed a pitch accent H+L* on the last phonological word of the initial DP and also on the last word of the utterance. A final low boundary tone L% was also found. The Subject Condition (SC) showed a pitch accent H+L* only on the last word of the utterance and a final low boundary tone L%. Concerning durational measurements, differently from the Topic Conditions (Figures 1 and 2), the last word of the initial DPs did not show lengthening of the nuclear and the post-nuclear syllables. Although the Long Subject DP Condition showed a pitch accent on the last word of the initial DP, there are no acoustic cues of lengthening or pause that characterize this DP as a single IP. The initial DPs of the Subject Conditions (LSC and SC) are within a phonological phrase (ip). Spectrograms of stimuli in Subject Conditions (Figures 3 and 4) are shown below. The red circles represent the DPs that were isolated to compose stimuli A and B in the task.

FIGURE 3 – Long Subject DP Condition (LSC): pitch track for item
The party portrait album was kept in the drawer

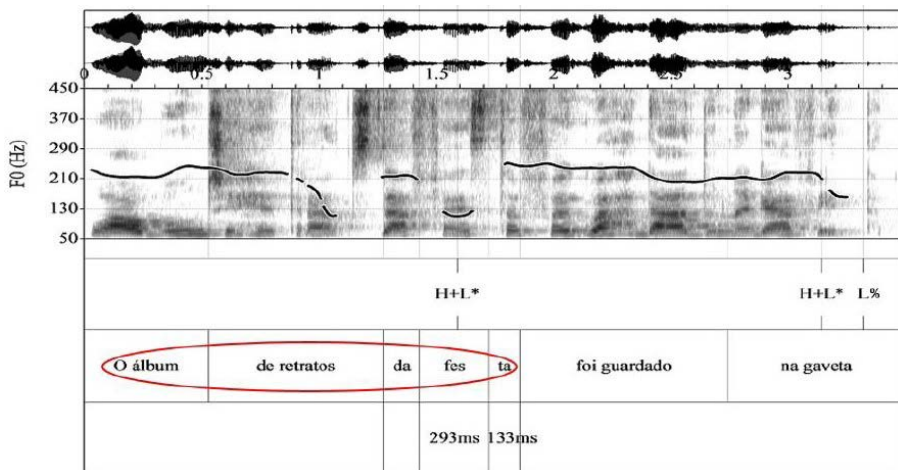
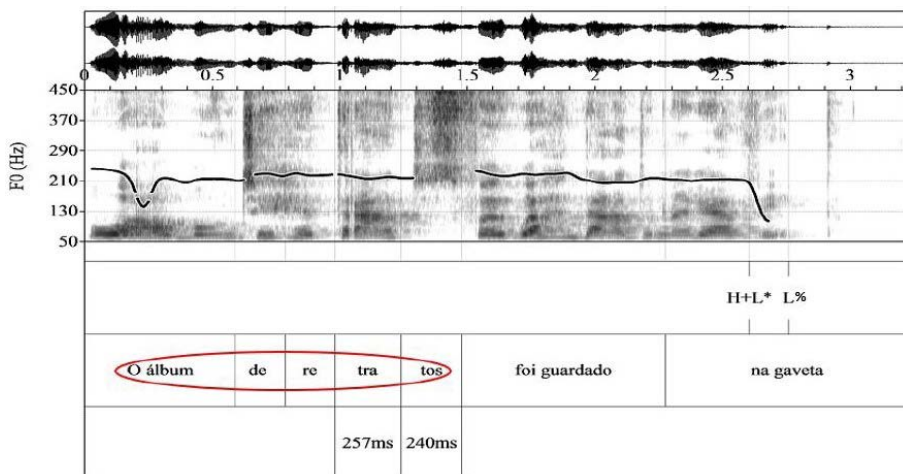


FIGURE 4 – Subject Condition (SC): pitch track for item
The portrait album was kept in the drawer



3.2 Procedures

This experiment was carried out on a personal laptop running DMDX software (FORSTER; FORSTER, 2002). Participants were seated at a desk in a quiet room in front of a laptop. They wore headphones to hear each experimental item and they also held a joystick, which was used to choose between stimuli A and B. The experimental items were counterbalanced so each participant heard an equal number of trials in each condition, in an individually-randomized order. The correct responses to the questions, A or B, were also counterbalanced.

Firstly, the sentence in one of the four conditions (stimulus X) was played through headphones. After that, the word SOUND A appeared on the left side of the screen and the initial DP was played in one of the prosodic versions, topic or subject. Subsequently, the word SOUND B appeared on the right side of the screen and the other DP was played in another prosodic version. After hearing stimuli A and B, participants read the following question on the screen: *Which stimulus is contained in the sentence? SOUND A or SOUND B?* The participants were to choose the DP (Sound A or Sound B) that matched acoustically the initial DP of the sentence (Stimulus X) that they had previously heard. Participants were instructed to press the button on the left of the joystick (marked with a

sticker with the letter A written on it) for the answer on the left side of the screen, or a button on the right (marked with a sticker with the letter B written on it) for the answer on the right side. The computer recorded response times and response choices. Each subject saw an equal number of items in each condition over the experiment in a Latin-square design. Each experimental session lasted between 15 and 20 minutes.

3.3 Participants

The participants were 24 native Brazilian Portuguese speakers (19 female and 6 male) who reported normal hearing and vision. The mean age of the sample was 33.3 years old. Subjects were high school students at the Educational Project for Young People and Adults (EJA). Some students were from John XXIII Application School and others were from Federal Institute of the Southeast of Minas Gerais (*Campus Juiz de Fora*). The participants signed a term of consent and volunteered to take part in the experiment. For task performance, subjects were equally divided into four groups.

3.4 Results and discussion

A table with the percentages of correct responses, incorrect responses and missed responses for each experimental condition is shown below:

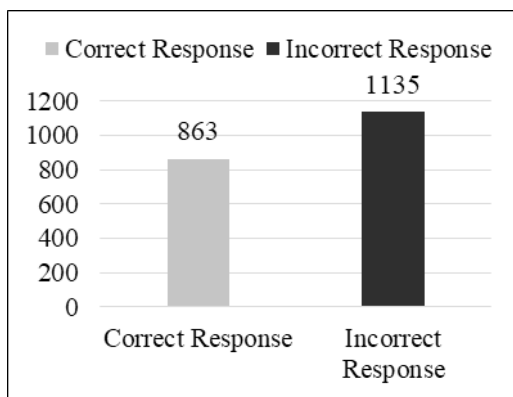
TABLE 2– Response Percentages of ABX task

Response Percentages			
Condition	Correct responses (%)	Incorrect responses (%)	Missed responses (%)
Long Topic DP Condition	70,8	21,9	7,3
Topic Condition	69,8	26,0	4,2
Long Subject DP Condition	77,1	18,8	4,2
Subject Condition	65,6	27,1	7,3
Total	70,8	23,4	5,7

The database indicate that DPs were correctly matched to stimulus X most of the time, approximately 70 % of accuracy. The rate of missed responses was disregarded, leaving out rates of correct responses and incorrect responses at 75% and 25%, respectively. The rates of correct responses and incorrect responses were submitted to a binomial non-parametric statistical test, which revealed that there was significant statistical difference between the two rates ($p < 0.001$).

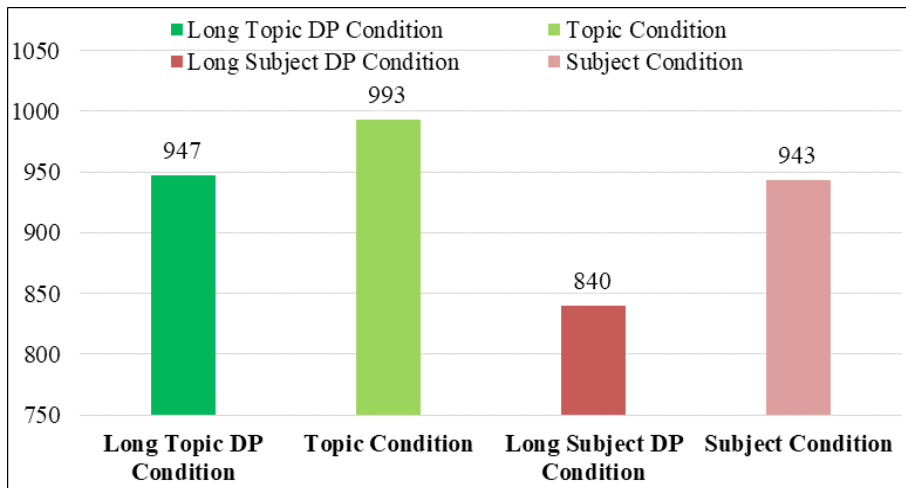
A graph of the average response times (RTs) subjects spent to choose a correct reponse or an incorrect reponse is included below. The results of Graph 1 indicate that participants spent more time when they had chosen an incorrect response.

GRAPH 1 – Response time averages (ms) by choice of Correct Response and Incorrect Response



RT averages were submitted to a t-Student statistical test for paired samples, which indicated a significant difference between RTs of correct reponses and RTs of incorrect responses: $t(360) = -57,456$; $p < 0.001$. RT averages by each condition were also analyzed – see Graph 2 below. ANOVA *post-hoc* Bonferroni did not reveal any statistical differences between the four conditions.

GRAPH 2 – Response time averages (ms) per condition



To summarize, the results of Table 1 indicate that hearers are able to perceive prosodic differences between topicalized DPs and non-topicalized subject DPs, since there was approximately 70% accuracy. The results of Graph 1, which showed slower RTs for choices of incorrect responses, suggest that participants did not respond at random. These slower RTs may be signaling that subjects had chosen an incorrect response because they were facing difficulties in auditory recognition. The results of Graph 2, which point to similarity of RTs per experimental condition, suggest that there was no condition that was more difficult to understand. Therefore, it is possible to conclude that hearers are able to perceive prosodic differences between topicalized DPs and non-topicalized DPs in subject position.

4 Experiment 2: Cross-modal naming with pictures

This experiment was designed to elicit topic-comment sentences and subject-predicate sentences in contexts created to favor the occurrence of such syntactic structures in speech. Cross-modal naming task is a type of on-line experiment that is used to measure processing at the point of syntactic disambiguation. In these tasks, participants listen to an auditory fragment followed by a visual target that is either an appropriate or an

inappropriate continuation of the sentence fragment. Subjects are required to name the visual target as quickly as they can and then use the target to complete the sentence. Completions are to ensure that participants are able to integrate the auditory fragment and the visual target, and to indicate the final structure and interpretation. Naming times are measured in order to reflect the easiness or the difficulty of integrating the visual target and the auditory fragment together into a sentence (TYLER; MARSLEN-WILSON, 1977; MARSLEN-WILSON *et al.*, 1992; KJEELGARD; SPEER, 1999; BLODGETT, 2004).

For the current research, the cross-modal naming task was adapted by using pictures instead of auditory materials. Participants visualized a picture that favored the speech production of an animate DP or an inanimate DP. Following the picture, they visualized the target word that favored either the construction of a topic-comment syntactic structure or the construction of a subject-predicate syntactic structure. Subjects were required to produce aloud the beginning of a sentence by integrating the picture and the visual target and then complete the rest of the sentence with some idea, so that the whole sentence was meaningful.

Animate DPs and inanimate DPs were chosen as objects of investigation because it is argued in the linguistics literature that there is a relation between animacy and agentivity. According to Lima Júnior and Côrrea (2015), speakers tend to place thematic roles of agent in the subject position. The role of agent is usually attributed to an animate constituent (FERREIRA, 1994). According to these authors, speakers tend to manifest a preference for active sentences with animate subject as opposed to passive ones, for example. Based on these studies, it is hypothesized that in the current task it will be easier for participants to create subject DPs in conditions that the picture favors an animate DP and it will be easier to create topic DPs in conditions that the picture favors an inanimate DP.

Therefore, this experiment was designed to achieve three goals: (i) to investigate whether in contexts that favor the production of subject-predicate structures and topic-comment structures participants are able to produce sentences consistent with such syntactic structures; (ii) second, to identify whether there is a default preference in speech for one of the two structures; (iii) to verify if animacy is a factor that influences the choice of the syntactic constructions.

4.1 Materials

Stimuli were constructed according to a design 2x2: (i) DP type favored by the picture: animate DP or inanimate DP; (ii) type of visual target word that follows the picture: subject pronoun or linking verb. This design allowed the construction of four conditions, which were named as: Animate Topic DP, Animate Subject DP, Inanimate Topic DP and Inanimate Subject DP. The Topic Conditions differed from the Subject Conditions just in relation to the type of visual target. For the Topic Conditions, the visual target word was the subject pronoun ‘he’ (*ele*), ‘she’ (*ela*) or ‘it’ (*ele/ela*). The type of subject pronoun that appeared after the picture – he, she or it – was chosen in order to match in genre to the element biased by the picture. These subject pronouns also allowed participants the possibility of using the initial DP as a referent in their sentences. For the Subject Conditions, the visual target word was the linking verb ‘was’ (*era, foi*).

Four stimuli for each condition were constructed, sixteen in total. An example of each condition is shown below:

- (8) **Animate Topic DP:** Picture (Animate DP) + Pronoun (‘she’ or ‘he’)



+ ELA (SHE)...

Possible DP: *A garota de bolsa vermelha...*

(The girl with the red purse...)

- (9) **Animate Subject DP:** Picture (Animate DP) + Verb ('was')



+ *ERA* (WAS)...

Possible DP: *O cachorro magro...*(The skinny dog...)

- (10) **Inanimate Topic DP:** Picture (Inanimate DP) + Pronoun ('it')



+ *ELE* (IT)...

Possible DP: *O álbum de retratos...*(The portrait album...)

- (11) **Inanimate Subject DP:** Picture (Inanimate DP) + Verb ('was')



+ *FOI* (WAS)...

Possible DP: *A parede da sala...*(The living room wall...)

In addition to the experimental stimuli, twelve other sentences were created. These sentences presented DPs or PPs that could be easily elicited through the pictures. Regarding the syntactic structure, the pictures were followed by a linking verb such as 'is' (*está/fica*), or by subject pronouns such as 'I' (*eu*) or 'you' (*você*). Among these twelve sentences, two of them were chosen to compose the training session. Some examples are shown below:

(12) **Sentence 5:** Picture + Verb



+ *FICA* (IS)...

Possible DP: *O tênis de couro...* (The leather shoe...)

(13) **Sentence 7:** Picture + Pronoun



+ *VOCÊ* (YOU)...

Possible PP: *Na padaria...* (At the bakery...)

4.2 Procedures

This experiment was carried out on a personal laptop running DMDX software. The computer recorded the sentences produced by participants and their RTs right from the beginning of utterance of the sentence created. The experimental items were counterbalanced so each participant visualized an equal number of trials in each condition, in an individually-randomized order.

Subjects were individually placed in a quiet room. They were seated at a desk in front of a laptop. The experimental session began with instructions. Participants were told to look at a picture on the screen of the laptop that was immediately followed by the presentation of the visual target word that could continue the sentence. The presentation of the picture lasted for 250ms and so did the presentation of the visual target word. After the presentation, subjects were asked to integrate the picture

and the target word in order to create the beginning of a sentence. They were also asked to complete the rest of the sentence with their ideas, but the sentence should make sense. When the participants had already formed a complete sentence, they should say it aloud for recording. Each subject saw an equal number of trials in each condition over the experiment in a Latin-square design. Each experimental session lasted between 15 to 20 minutes.

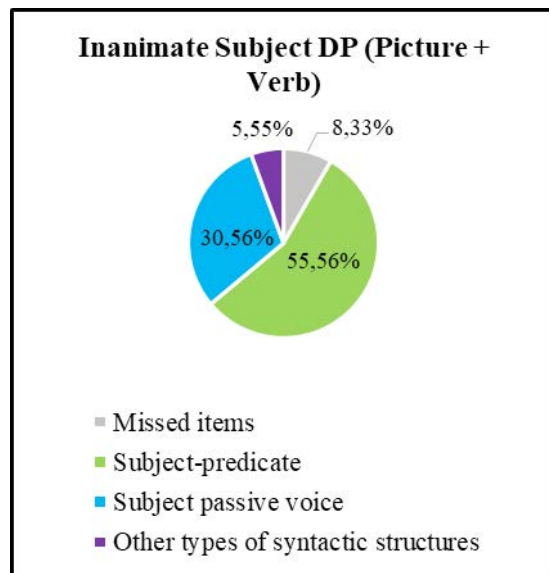
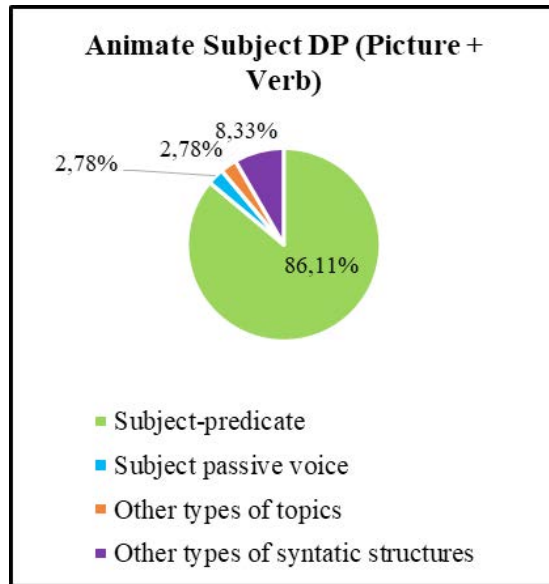
4.3 Participants

The participants were 18 native Brazilian Portuguese speakers (11 female and 7 male) who reported normal hearing and vision. The mean age of the sample was 19 years old. Subjects were undergraduate students at Federal University of Juiz de Fora. The participants signed a term of consent and volunteered to take part in the experiment. For task performance, subjects were equally divided into two groups.

4.4 Results and discussion

Graph 3 presents the types of syntactic structures of the sentences produced by participants in conditions Animate Subject DP and Inanimate Subject DP. It is worth mentioning that in the category ‘other types of topics’ we grouped sentences that presented topic DPs with syntactic function of adjuncts or adverbs. On the other hand, in the category ‘other types of syntactic structures’ we grouped relative sentences, conjoined sentences, exclamatives and questions.

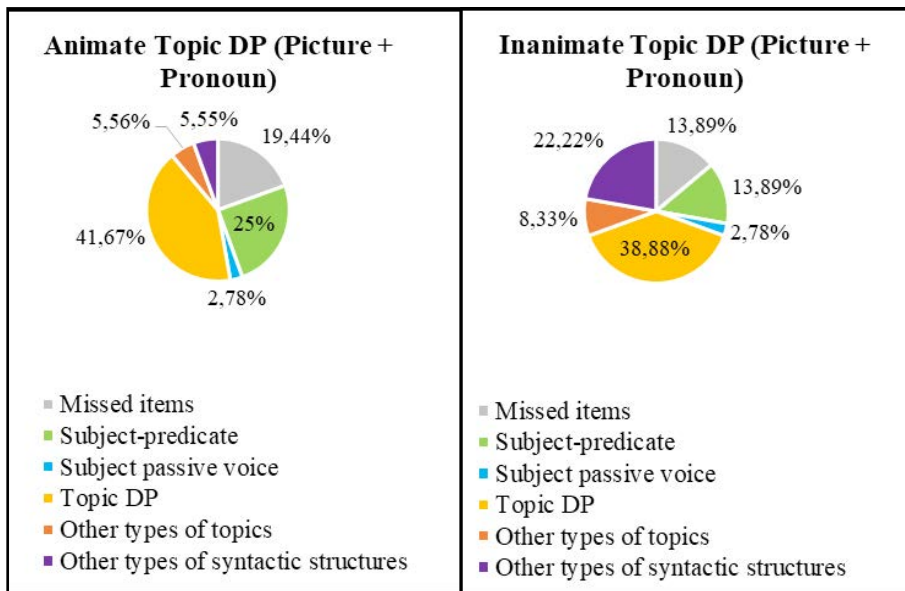
GRAPH 3 – Syntactic structure of the sentences produced by participants for the conditions of Subject DP



Condition Animate Subject DP received more subject-predicate responses (around 86%) as expected. It is also possible to notice that this condition did not present missed responses. Regarding condition Inanimate Subject DP, there were more subject-predicate responses (around 55%) as it was expected. However, there was a high percentage of sentences with subject passive voice. It seems that the factor animacy influenced participants' syntactic choices. This result is in line with what is claimed by Lima Júnior and Côrrea (2015) and Ferreira (1994), speakers tend to place in the subject position an animate constituent with thematic role of agent. Thus, participants may have found it difficult to create a subject-predicate sentence in active voice with an inanimate DP.

Graph 4 presents the types of syntactic structures of the sentences produced by participants in conditions Animate Topic DP and Inanimate Topic DP. In the category 'Topic DP' we grouped the productions in which there was topicalization of the subject of the comment sentence and the productions in which there was the topicalization of the object of the comment sentence.

GRAPH 4 – Syntactic structure of the sentences produced by participants for the conditions of Topic DP



The conditions Animate Topic DP and Inanimate Topic DP presented respectively 41.67% and 38.88% of responses with topic-comment structure. Overall, the results indicate that the context was able to increase the responses with topic-comment structure, but this syntactic structure was not unanimously chosen, since different types of structure occurred. Furthermore, the Topic Conditions presented highest rates of missed responses, the condition Animate Topic DP was the condition that presented the highest rate of missed responses, around 19%. It seems that the factor animacy was also influential, since the condition Animate Topic DP presented 32.1% of responses with subject-predicate structure. Participants may have found difficult to create a topic-comment sentence and decided to ignore the subject pronoun and replace it with a verb, or to put it in another position in the sentence. This result is also in line with the claims made by Lima Júnior and Côrrea (2015) and Ferreira (1994). One result was puzzling though; it was expected that pictures in condition Inanimate Topic DP would facilitate production of topic DPs, but this did not happen, since there are more sentences with topic-comment structure in Animate Topic DP. A possible explanation is that when participants visualized a personal pronoun after the picture, they promptly associated it with the subject of the sentence, which was the animate DP biased by the picture. Thus, they used the subject pronoun to refer to the subject of the sentence.

If both graphs are to be compared, one interesting result is the fact that in Subject Conditions no productions with topic-comment structure are seen, whereas in Topic Conditions, productions with the subject-predicate structure occurred. These results suggest that speakers seem to prefer the subject-predicate structure as the default syntactic structure in BP. Participants only produced sentences with topic-comment structure when there was a bias favoring the occurrence of such structure, that is, when the visual target word was a subject pronoun.

Here are some examples of the sentences produced by participants:

(14) Condition **Animate Topic DP**

Production of Topic DP sentence type, by subject S2INFO7:

A modelo, ela é linda. (The model, she is gorgeous)

(15) Condition **Inanimate Topic DP**

Production of Topic DP sentence type, by subject S2INFO5:

A mochila vermelha, ela usou para ir ao trabalho. (The red backpack, she wore [it] to go to work)

(16) Condition **Animate Subject DP**

Production of Subject-Predicate sentence type, by subject S1INFO2:

O cachorro era de rua. (The dog was living on the street)

(17) Condition **Inanimate Subject DP**

Production of Subject-Predicate sentence type, by subject S1INFO9:

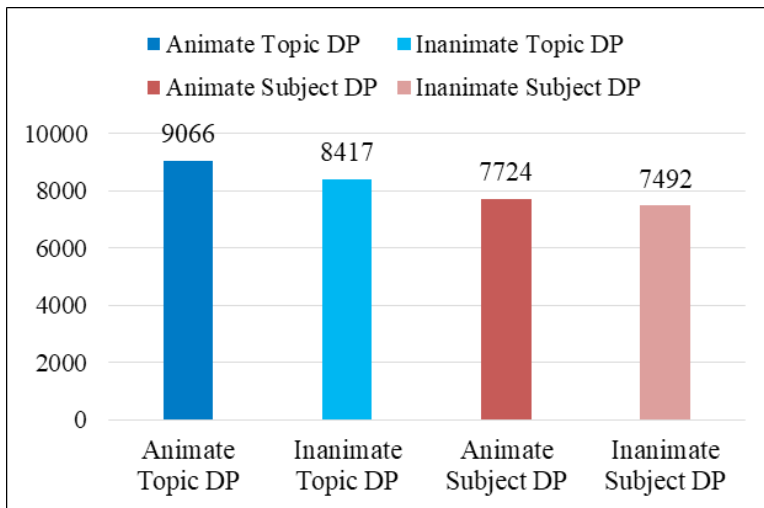
O filme foi excelente. (The movie was great)

Production of Subject Passive Voice sentence type, by subject S2INFO4:

A foto foi revelada. (The photo was developed)

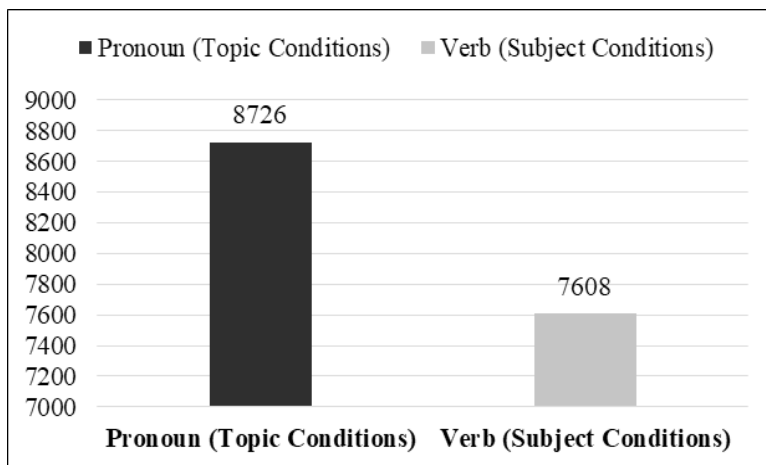
The average RTs participants spent after they had visualized the picture and the visual target word to start saying their sentences aloud were also analyzed. Graph 5 shows the average RTs of each condition. Overall, this graph shows longer RTs to create sentences in Topic Conditions as opposed to Subject Conditions. We submitted average RTs of the four conditions to the ANOVA *post-hoc* Bonferroni test and no significant differences between them were found.

GRAPH 5 – Average RTs (ms) of the four conditions



Therefore, in Graph 6 the average RTs were grouped according to the visual target word: the category Pronoun contained the conditions Animate Topic DP and Inanimate Topic DP; whereas the category Verb contained the conditions Animate Subject DP and Inanimate Subject DP.

GRAPH 6 –Average RTs (ms) of conditions classified by visual target word type



The data concerning the average RTs were analyzed by means of ANOVA. Within-subjects ANOVA and within-items ANOVA with two conditions of animacy (animate x inanimate) and two conditions of visual target word (pronoun x verb) were conducted. In the within-subjects analysis there was no main effect of animacy type $F(1,127) = 1.080$, $p = 0.301$, but there was a main effect of visual target word type $F(1,127) = 6,446$, $p = 0.012$. There was no main effect of interaction between animacy type and visual target word type $F(1,127) = 0.026$, $p = 0.873$. Within-items analysis showed similar results, there was no main effect of conditions animacy type $F(1,127) = 2.846$, $p = 0.094$, but there was a main effect of visual target word type $F(1,127) = 5,748$, $p = 0.018$. There was no main effect of interaction between the animacy type and the visual target word type $F(1,127) = 0.031$, $p = 0.861$.

In summary, the results indicate that participants faced more difficulty to create sentences in Topic Conditions, in which the picture was followed by a subject pronoun, than to create sentences in Subject Conditions, in which the picture was followed by a verb. One evidence of this difficulty was the longer RTs found in Topic Conditions. The rates of production with the target syntactic structure also point to this difficulty, since the production rates in Topic Conditions, 41.67% in Animate Topic DP and 38.88% in Inanimate Topic DP, were lower than the rates of production found in Subject Conditions, 86.11% in Animate Subject DP and 55.56% in Inanimate Subject DP. This difficulty may have been due to the fact that the topic-comment structures are considered as specific constructions in BP and, thus, they could be more dependent on the discursive context. Subject-predicate structures, on the other hand, may have been easier to produce because they are more recurrent in speech. Therefore, although participants had produced more topic-comment sentences when the experimental conditions biased the occurrence of this structure, productions with subject-predicate structure were shown to be preferred. In BP, the subject-predicate structure seems to be the default.

5 Experiment 3: Self-paced listening and reading

This experiment was designed to verify whether prosodic characteristics of a DP in topic position or in non-topicalized subject position are informative for hearers to distinguish between these two syntactic categories. It also aims to verify whether participants are able

to perceive when there is a possible mismatch between prosody of the initial DP and the word (name or verb) that comes next in the sentence.

According to Rayner and Clifton (2002), self-paced task is an online experiment that allows researchers to verify how long it takes a subject to read or listen to a particular input. The experimenter is able to control the amount of input that participants can read or listen to (word-by-word, phrase-by-phrase), depending on the study object under investigation. Participants determine the rate at which the material is presented. The task involves pressing a particular button to read or listen to segment-by-segment. When subjects have understood the segment, they push a button and the next segment is presented. After the presentation of the whole sentence, a question appears on the screen in order to verify participants' understanding of the sentence and also to keep their attention on the task. The program in which the task is carried out records the time to read or listen to each segment. The reading task can present a cumulative design or a non-cumulative design. In a cumulative design, words that have been revealed are kept on screen until the end of the whole presentation, whereas in a non-cumulative design words that have already been read disappear when the participant presses the button to reveal the next segment. This methodological difference also depends on the interests of the research. According to Garrod (2006), this technique has been widely used to investigate syntactic analysis, speech comprehension processes and the resolution of anaphora especially. Self-paced is advantageous because it gives a good indication of when the participant encounters some difficulty in comprehension.

For the current research, a self-paced task that combined listening and reading was designed. The first segment was auditory and the other segments were written. That is, participants listened to the first segment that contained the initial DP with prosody of topic or prosody of subject. After listening to that segment, they pressed the button to read the other segments that gave continuity to the sentence. After the whole presentation of the experimental item, they pressed the button to read the comprehension question. The segment that appeared after the auditory stimulus (the topicalized DP or the subject DP) was considered the critical segment of the sentences, since it was the point of a possible mismatch between prosody and syntax. This type of design was chosen due to the possibility of controlling the size of segments – a factor that could influence the response times – and minimizing coarticulation effects

between the DP and the word that followed it in Subject Conditions. Sentences in Topic Conditions did not show coarticulation effects, since there was a pause between the initial DP and the following word. However, the sentences in Subject Conditions showed such effect due to the lack of a pause. It was necessary to record these sentences with the same linguistic input after the initial DP in order to neutralize the coarticulation. After the initial DP a verb that was initiated by a voiceless plosive consonant was revealed, which allowed a micropause to occur. The presence of this micropause facilitated the isolation of the initial DPs in Praat.

5.1 Materials

Stimuli were constructed according to a design 2x2x2: (i) type of syntactic structure: topic-comment or subject-predicate; (ii) initial DP size: seven-syllable DP or four-syllable DP; (iii) congruence between prosody of the initial DP and target word that gives continuity to the syntactic structure: congruent or incongruent. This design allowed the construction of eight conditions. Both the Topic Conditions and the Subject Conditions were initiated by the same type of DPs, which were different just in regard to prosody. Short DPs contained four syllables, whereas long DPs contained seven syllables. All congruent topic-comment sentences contained an initial DP, a noun, a direct verb and a PP. All congruent subject-predicate sentences contained an initial DP, a direct verb, an object and a PP. With respect to incongruence, the Incongruent Topic Conditions contained the initial DP with topic prosody and the syntactic structure of the subject-predicate sentences. That is, after the initial topic DP a verb appeared, which was incongruent with topic prosody. The Incongruent Subject Conditions contained the initial DP with baseline subject prosody and the syntactic structure of topic-comment sentences. That is, after the initial subject DP a noun, which was incongruent with subject prosody, appeared.

Therefore, this self-paced study contained eight conditions. The sentences were broken up into four segments, as shown by the slashes:

(18) **Condition Short Topic DP – Congruent**

L+H* H% H+L* L%

O gerente/ o dono / demitiu / sem motivo.
 The manager/ the boss / fired (him) / without any reasons.

(19) **Condition Short Topic DP – Incongruent**

L+H* H% H+L* L%

O gerente/ delegou / tarefas / ao garçom.
 The manager / delegated / duties / to the waiter.

(20) **Condition Short Subject DP – Congruent**

H+L* L%

O gerente / delegou / tarefas / ao garçom.
 The manager / delegated / duties / to the waiter.

(21) **Condition Short Subject DP – Incongruent**

H+L* L%

O gerente / o dono / demitiu / sem motivo.
 The manager / the boss / fired (him) / without any reasons.

(22) **Condition Long Topic DP – Congruent**

LH L+H* H% H+L* L%

O gerente do bistrô / o dono / demitiu / sem motivo.
 The bistro manager / the boss / fired (him) / without any reasons.

(23) **Condition Long Topic DP – Incongruent**

LH L+H* H% H+L* L%

O gerente do bistrô / delegou / tarefas / ao garçom.
 The bistro manager / delegated / duties / to the waiter.

(24) **Condition Long Subject DP – Congruent**

H+L* L%

*O gerente do bistrô/ delegou / tarefas / ao garçom.*The bistro manager / delegated / duties / to the waiter.(25) **Condition Long Subject DP – Incongruent**

H+L* L%

*O gerente do bistrô / o dono / demitiu / sem motivo.*The bistro manager / the boss / fired (him) / without any reasons.

Ninety-six stimuli were elaborated in total, that is, there were twelve sentences for each condition. The sentences were recorded by the same native BP speaker who recorded stimuli for Experiment 1. After the recording, the software Praat was used to isolate the initial DPs of the stimuli. There was a manipulated pause of about 100ms after the initial DPs in Topic Conditions. The initial DPs in Subject Conditions did not present pauses.

In addition to experimental items, thirty sentences were elaborated. Some initial DPs of these sentences were recorded with baseline prosody, while other initial DPs were recorded with focus. Four sentences, among these thirty-one, were chosen to appear in the practice round.

5.2 Procedures

This experiment was conducted using DMDx software on a personal laptop. Subjects were individually taken to a quiet room. They were seated at a desk in front of the laptop. Each experimental session began with instructions followed by a short practice round to familiarize them with the task. In the practice, they were exposed to four unrelated sentences and they answered a comprehension question after each sentence. Each trial began when a participant pressed a particular button of the joystick. The auditory segment was played through headphones. They were to press the button of the joystick again when they had heard and understood the segment. The following segments were all written. Thus, they pressed the button again to read the second segment of the sentence, pressed it again to see the third segment, and pressed it again

when they were done reading the sentence. They were instructed to read at a comfortable pace that allowed them to comprehend the sentences. After the presentation of each item, a yes/no comprehension question appeared on the screen. They also pressed one of the joystick buttons to answer these questions. DMDx recorded response times (RTs) of the segments as well as the answers to the comprehension questions and RTs to answer them. The items appeared in individually randomized order such that no consecutive trials were of the same type. Each subject saw an equal number of items in each condition over the experiment in a Latin-square design. Each session lasted between 15 to 20 minutes.

5.3 Participants

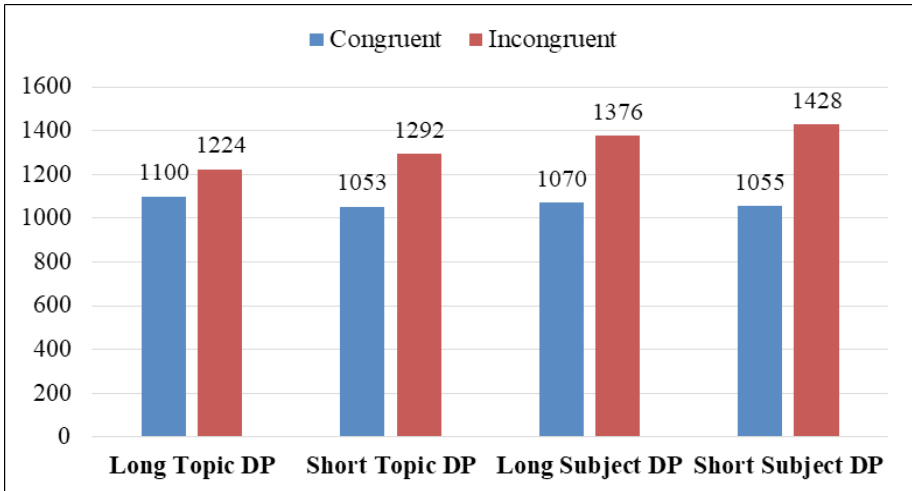
The participants were 24 native Portuguese-speaking adults (19 female and 5 male) who reported normal hearing and vision. The mean age of the sample was 23,3 years old. Subjects were undergraduate students at Federal University of Juiz de Fora. They all signed a term of consent and volunteered to take part in the experiment. For task performance, subjects were equally divided into four groups.

5.4 Results and Discussion

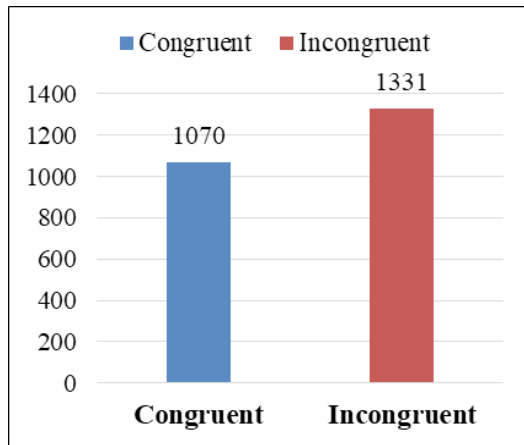
The software DMDx recorded RTs for the four segments of each sentence. However, just the second segments of the sentences were analyzed because they were the critical ones. In congruent conditions, the second segment should indicate that prosody of the initial DP matched the target word, a noun in Topic Conditions and a verb in Subject Conditions. In incongruent conditions, the second segment should indicate that the prosody of the initial DP did not match the target word, a noun in Subject Conditions and a verb in Topic Conditions. For the analysis, any RTs under 200 ms or over 3500 ms were disregarded.

Graph 7 shows average RTs of the second segment of each condition and Graph 8 shows average RTs of the second segment considering the two large groups, congruent conditions and incongruent conditions. Both graphs indicate that incongruent conditions presented slower reading times in comparison to congruent conditions. In Graph 7, it is possible to notice that the Incongruent Subject Conditions presented the greatest RTs.

GRAPH 7 – Average RTs (ms) of the critical segment of each experimental condition



GRAPH 8 – Average RTs (ms) of critical segment: group of congruent conditions and group of incongruent conditions



Average RTs of congruent conditions and incongruent conditions were submitted to within-subjects ANOVA and within-items ANOVA with 2x2x2 design (type of DP size: long DP and short DP x two types of syntactic structure: topic and subject x two types of prosody: congruent and incongruent). Within-subjects analysis did not reveal main effects

of DP type $F(1,569) = 0.085$, $p = 0.770$ or syntactic type $F(1,569) = 3,262$, $p = 0.071$. However, there was main effect of type of prosody $F(1,569) = 48,186$, $p < 0.001$. The analysis also revealed that there was interaction effect between the syntactic structure and prosody $F(1,569) = 6,913$, $p = 0.009$. Within-items analysis presented similar results, there was no main effect of DP size $F(1,569) = 0.095$, $p = 0.758$ or syntactic structure $F(1,569) = 3.257$, $p = 0.072$, but there was main effect of prosody $F(1,569) = 48,150$, $p < 0.001$. The analysis also revealed that there was only interaction effect between the syntactic structure and the prosody $F(1,569) = 6,907$, $p = 0.009$. ANOVA *post-hoc* Bonferroni was also conducted to compare incongruent conditions to their congruent versions. The analysis indicated that there were no significant differences between the Incongruent Topic Conditions and the Congruent Topic Conditions. However, there were significant differences between the Incongruent Subject Conditions and the Congruent Subject Conditions.

Overall, the results indicate that subjects recognized the prosody-syntax incongruence only in Subject Conditions, due to significant statistical difference of RTs encountered in the ANOVA Bonferroni test between the Congruent Subject Conditions (Long Subject DP: 1070ms; Short Subject DP: 1055ms) and the Incongruent Subject Conditions (Long Subject DP: 1376ms; Short Subject DP: 1428ms). It seems that completion was contrary to participants' expectation. That is, when they heard the initial DP with a baseline prosody of subject, they might have expected a verb to continue the sentence, but a noun appeared instead. This counter-expectation was manifested in reading latencies. Regarding the Topic Conditions, although average RTs indicated that participants had spent more time to read the critical segment in the incongruent conditions (Long Topic DP: 1224 ms; Short Topic DP: 1292 ms) than to read the segment of congruent conditions (Long Topic DP: 1100ms; Short Topic DP: 1053ms), the ANOVA Bonferroni did not indicate any statistical differences between them. One possible explanation for this result may be that participants perceived the initial DP as a focused subject, and so there was no counter-expectation when they visualized the verb because this condition is not totally incongruent. This condition could be interpreted as congruent in a discursive situation where prosodic strengthening of the initial DP in the subject position was required. One interesting result is the fact that RTs of critical segments in congruent conditions of Topic and Subject were similar. These data seem to suggest

that when the topicalized constituent receives proper prosodic cues, participants process both structures in a similar manner. A possible explanation for this result may be that because the topic is a marked structure in BP it needs to receive proper cues, such as prosodic ones, in order to be recognized promptly as other common syntactic structures in BP are, like subject-predicate structures for instance.

In summary, the results of this experiment indicate that subjects were able to identify prosodic cues present in the DPs; additionally they also used these characteristics in the processing of topic-comment and subject-predicate sentences. The results also show that prosody is an important component that has psychological reality in linguistic processing.

6 Conclusions

This research aimed to explore the role of prosody in the processes of comprehension and production of topic-comment and subject-predicate structures in BP. Three experimental tasks were carried out: a perception task with ABX technique, a production task with Cross-modal naming technique and a comprehension task with Self-paced listening and reading techniques. The perception/comprehension tasks allow us to conclude that hearers are able to recognize the prosodic differences between topic-comment sentences and subject-predicate sentences, and these prosodic features are informative enough for them to differentiate between these two syntactic structures during processing. The production task allows us to conclude that in contexts favorable to the occurrence of topic-comment and subject-predicate structures, speakers are able to produce sentences consistent with such syntactic structures.

In summary, the results of the three experiments together allow us to conclude that prosody is an important component in the processing of topic-comment and subject-predicate structures since speakers use the prosodic cues to process these structures. We also found out that topic-comment structures are processed and understood by speakers both syntactically and prosodically. Finally, we do not have evidences to suggest a process of changing in the typological status of BP to a type (iii) language in Li and Thompson's typology (1976), a language with prominence of both subjects and topics, since our results suggest that the subject-predicate syntactic structure remains the default in BP.

Acknowledgements

We would like to thank all the students that volunteered to take part in the experiments. This research was funded and supported by the Foundation for Research Support of Minas Gerais (FAPEMIG) and Federal University of Juiz de Fora (UFJF).

Authors' contributions

Silva, A. C. O. and Fonseca, A. A. conceived the presented idea. Silva, A. C. O. developed the theoretical framework and Fonseca, A. A. provided feedback. Silva, A. C. O. and Fonseca, A. A. planned and carried out the experiments. Silva, A. C. O. collected data results and Fonseca, A. A. performed the statistical analysis. Silva, A. C. O. and Fonseca, A. A. discussed the interpretation of the results. Silva, A. C. O. wrote the manuscript with support from Fonseca, A. A.

References

- BECKMAN, M. E.; PIERREHUMBERT, J. B. Intonational structure in Japanese and English. *Phonology*, Cambridge, v. 3, p. 255-309, 1986. Doi: <https://doi.org/10.1017/S095267570000066X>
- BLODGETT, A. Functions of Intonation Boundaries during Spoken Language Comprehension in English. In: INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE - INTERSPEECH, 8th, 2004, Jeju Island. *Processing...* Jeju Island: International Convention Center Jeju, 2004. p. 2997-3000.
- BOERSMA, P.; WEENICK, D. *PRAAT: doing phonetics by computer* (version: 5.3.22), 2008. <<http://www.praat.org/>>.
- BOLEY, J.; LESTER, M. Statistical Analysis of ABX Results Using Signal Detection Theory. *Journal of the Audio Engineering Society*, New York, p. 1-7, 2009.
- CALLOU, D.; MORAES, J.; LEITE, Y.; KATO, M. A.; OLIVEIRA, C. T.; COSTA, E.; ORSINI, M.; RODRIGUES, V. Topicalização e Deslocamento à esquerda: Sintaxe e Prosódia. In: CASTILHO, A. T. (Org). *Gramática do Português Falado Volume III: As Abordagens*. Campinas: Editora da Unicamp, 1993. p. 315-360.

CARLSON, K.; CLIFTON JR., C.; FRAZIER, L. Prosodic boundaries in adjunct attachment. *Journal of Memory & Language*, Elsevier, v. 45, n. 1, p. 58-81, 2001. Doi: <https://doi.org/10.1006/jmla.2000.2762>

CLIFTON JR., C.; CARLSON, K.; FRAZIER, L. Informative prosodic boundaries. *Language and Speech*, Sage Journals, v. 45, p. 87-114, 2002.

FERREIRA, F. Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language*, Elsevier, v. 33, p. 715-736, 1994. Doi: <https://doi.org/10.1006/jmla.1994.1034>

FONSECA, A. A. *A prosódia no parsing: evidências experimentais do acesso à informação prosódica no input linguístico*. 2012. 208 f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2012.

FORSTER, J.; FORSTER, K. *DMDX Display Software*, 2002. <<http://www.u.arizona.edu/~kforster/dmdx/dmdx.htm>>.

FRAZIER, L.; CLIFTON JR., C.; CARLSON, K. Don't break, or do: prosodic boundary preferences. *Lingua*, Elsevier, v. 114, p. 3-27, 2003. Doi: [https://doi.org/10.1016/S0024-3841\(03\)00044-5](https://doi.org/10.1016/S0024-3841(03)00044-5)

FROTA, S.; CRUZ, M.; SVARTMAN, F.; COLLISCHONN, G.; FONSECA, A.; SERRA, C.; OLIVEIRA, P.; VIGÁRIO, M. Intonational variation in Portuguese: European and Brazilian varieties. In: FROTA, S.; PRIETO, P. (Ed.). *Intonation in Romance*. Oxford: Oxford University Press, 2015. p. 235-283. Doi: <https://doi.org/10.1093/acprof:oso/9780199685332.003.0007>

GARROD, S. Psycholinguistic Research Methods. In: BROWN, K. (Ed.). *Encyclopedia of Language & Linguistics*. Oxford: Elsevier, 2006. p. 251-257. Doi: <https://doi.org/10.1016/B0-08-044854-2/04155-9>

GUSSENHOVEN, C.; JACOBS, H. *Understanding Phonology*. 3 ed. London: Hodder Education, 2011.

KENEDY, E. O status tipológico das construções de tópico no Português Brasileiro: uma abordagem experimental. *Revista da ABRALIN*, Curitiba, v. 13, n. 2, p. 151-183, jun./dez. 2014.

KENEDY, E. Tópicos e Sujeitos no PB: uma abordagem experimental. *Revista da Anpoll*, Florianópolis, v.1, n. 31, p. 69-88, 2011.

KJELGAARD, M. M.; SPEER, S. R. Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, Elsevier, v. 40, p. 153-194, 1999. Doi: <https://doi.org/10.1006/jmla.1998.2620>

LI, C.; THOMPSON, S. Subject and Topic: A New Typology of Language. In: LI, C. (Ed.). *Subject and Topic*. New York, Academic Press, 1976. p. 457-489.

LIMA JÚNIOR, J. C.; CÔRREA, L. M. S. A natureza do custo computacional na compreensão de passivas: um estudo experimental com adultos. *Letras de Hoje*, Porto Alegre, v. 59, n. 1, p. 91-101, jan./mar. 2015.

MARSLEM-WILSON, M. D.; TYLER, L. K.; WARREN, P.; GRENIER, P.; LEE, C. S. Prosodic effects in minimal attachment. *Quarterly Journal of Experimental Psychology*, Saage Journals, v. 45A, n. 1, p. 73-87, 1992.

MORAES, J.; ORSINI, M. T. Análise prosódica das construções de tópico no português do Brasil: estudo preliminar. *Letras Hoje*, Porto Alegre, v. 38, n. 4, p. 261-272, dez. 2003.

NESPOR, M.; VOGEL, I. *Prosodic Phonology: with a new foreword*. Berlim: Mouton de Gruyter, 2007. Doi: <https://doi.org/10.1515/9783110977790>

ORSINI, M. T. *As construções de tópico no português do Brasil: uma análise sintático-discursiva e prosódica*. 2003. 197 f. Tese (Doutorado em Língua Portuguesa) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2003.

PERINI, M. A. Topicalização. In: _____. *Gramática do Português Brasileiro*. São Paulo: Parábola Editorial, 2010. p. 331-335.

PIERREHUMBERT, J. *The Phonology and Phonetics of English Intonation*. 1980. 401 f. Thesis (Doctor of Philosophy) – Massachusetts Institute of Technology, Cambridge, 1980.

PONTES, E. *O tópico no Português do Brasil*. Campinas: Editora Pontes, 1987.

RAYNER, K.; CLIFTON JR, C. Language processing. In: MEDIN, D. (Ed.). *Stevens' Handbook of Experimental Psychology: Memory and Cognitive Processes*. 3. ed. New York: John Wiley and Sons; Copyright John Wiley & Sons. 2002. v. 2. p. 261-316. Doi: <https://doi.org/10.1002/0471214426.pas0207>

ROSS, J. R. *Constraints on Variables in Syntax*. 1967. 501 f. Thesis (Doctor of Philosophy) – Massachusetts Institute of Technology, Cambridge, 1967.

SILVA, A. C. O. *Processamento prosódico na compreensão e produção de estruturas de tópico e sujeito no Português Brasileiro*. 2017. 157f. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade Federal de Juiz de Fora, Juiz de Fora, 2017.

SILVA, C. G. C. *A interface prosódia-sintaxe na produção e no processamento de estrutura de tópico e de SVO*. 2015. 176 f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal de Juiz de Fora, Juiz de Fora, 2015.

TYLER, L. K.; MARSLER-WILSON, W. D. The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior*, New York, v. 16, p. 683-692, 1977. Doi: [https://doi.org/10.1016/S0022-5371\(77\)80027-3](https://doi.org/10.1016/S0022-5371(77)80027-3)



Complex Illocutive Units in L-AcT: An Analysis of Non-Terminal Prosodic Breaks of Bound and Multiple Comments

Unidades Illocucionárias Complexas na L-AcT: uma análise de quebras prosódicas não-terminais em Comentários Ligados e Comentários Múltiplos

Alessandro Panunzi

University of Florence, Florence, Italy

alessandro.panunzi@unifi.it

Valentina Saccone

University of Basel, University of Florence, Basel/Florence, Switzerland/Italy

valentina.saccone@unibas.ch

Abstract: This work presents a pilot study for a prosodic analysis of two different spoken structures in spoken Italian within the theoretical framework of the Language into Act Theory (L-AcT): (i) chains of two or more Bound Comments (COB) that do not form a compositional informative and prosodic unit; (ii) compositional Information Units formed by two or more Multiple Comments (CMM), linked together by a conventional prosodic model that implements specific meta-illocutive structures. This work analyzes COBs and CMMs from the DB-IPIC Italian Minicorpus. Different prosodic cues are taken into account: f_0 reset, pauses, final lengthening, intensity lowering and initial rush. The distinctive feature for COBs is a flat trend of f_0 before the boundary, with a low number of f_0 reset, while CMMs vary between different f_0 shapes. Vowel elongation and a no rushing speech rate cooperate in perceiving the prolongation of one COB into another. Initial rush is a characteristic feature of CMMs, while the lengthening of the last vowel of the unit is easier to find at the end of a COB than in a CMM.

Keywords: prosody; spontaneous speech segmentation; non-terminal breaks; L-AcT.

Resumo: Este trabalho apresenta um estudo piloto sobre uma análise prosódica de duas estruturas distintas em italiano falado, sob a perspectiva da Teoria da Língua em Ato (L-AcT): (i) cadeiras de dois ou mais Comentários Ligados (COB) que não formam

uma unidade informacional e prosódica composicional; (ii) unidades informacionais composicionais formadas por dois ou mais Comentários Múltiplos (CMM), ligados entre si por um modelo prosódico convencional que implementa estruturas metalocutivas específicas. Os COBs e CMMs analisados foram extraídos do *minicorpus* italiano disponível no DB-IPIC. Diferentes aspectos prosódicos são levados em conta: *reset* de f_0 , pausas, alongamento final, abaixamento de intensidade e *rush* inicial. O traço distintivo para os COBs é uma tendência a achatamento de f_0 antes da fronteira, com um baixo número de *reset* de f_0 , enquanto os CMMs variam entre diferentes formatos de f_0 . Alongamento de vogal e uma velocidade de fala sem *rushing* cooperam na percepção do prolongamento de um COB naquele que o segue. O *rush* inicial é um traço característico dos CMMs, enquanto o alongamento da última vogal da unidade é mais fácil de encontrar ao final de um COB do que de um CMM.

Palavras-chave: prosódia; segmentação da fala espontânea; quebras não-terminais; L-Act

Submitted on March 31st, 2018

Accepted on July 6th, 2018

1 Introduction

This work presents a description and an analysis of prosodic breaks in spontaneous spoken Italian, starting from a selection of examples included in the DB-IPIC resource (PANUNZI; GREGORI, 2012).¹ DB-IPIC is a linguistic database developed for the study of information structure strategies and their comparison in different languages. This resource includes the informal part of the Italian C-ORAL-ROM spoken corpus (CRESTI, MONEGLIA, 2005) and three Minicorpora of Italian, Brazilian Portuguese (from C-ORAL-BRASIL corpus; RASO; MELLO, 2012) and Spanish (from Cor-DiAL corpus; NICOLAS MARTINEZ, 2012), each one with the same size and design.

The analysis presented in this paper is a pilot corpus-based study, which aims at describing the formal differences between different types of non-terminal breaks co-occurring with two specific Information Units, as they are defined in the theoretical framework of Language into Act Theory (L-AcT; CRESTI, 2000; MONEGLIA; RASO, 2014). More

¹ Freely available online at <http://www.lablita.it/app/dbipic/>

specifically, this work deals with the prosodic and formal features of the tone units corresponding to Bound Comments and Multiple Comments as described below, and delineates a base for future prosodic studies on this matter.

We analyzed a sample including a total of 37 non-terminal prosodic breaks taken from 13 different recording sessions and different speakers, with the purpose of bringing out segmentation issues through formal acoustic parameters. The objects of our analysis were prosodic acoustic parameters on both sides of the tonal breaks. In this paper, on one hand, we aim to delineate typical ending features of Bound Comments and Multiple Comments, in order to simplify and help recognizing these units in speech flow. On the other hand, this work aims to individuate possible prosodic marks on the beginning of the new unit, whatever it might be, thus analyzing prosodic patterns just after the signaled break. In order to evaluate them, we used the Praat software (BOERSMA; WEENINK, 2005).

Section 2 presents an introduction of the theoretical framework, and Section 3 deepens the nature and characteristics of the Information unit treated in the analysis. In Section 4 we present the examples extracted from the corpus. Section 5 introduces the prosodic parameters used for the analysis, that it is reported in detail in Section 6.

2 Language into Act Theory

2.1 Theoretical foundations

Language into Act Theory originates from Speech Act Theory (AUSTIN, 1962). It is based on the observation of a systematic correspondence between pragmatic and prosodic units in speech, empirically verified through observation and analysis of tonal contours. This correlation extends on two hierarchical levels, each one linking the formal level of prosodic realization with the functional plane of pragmatic values. The superordinate level deals with the correlation between Speech Act production and terminal prosodic profiles, namely the *illocutionary principle*. The lower level looks at the isomorphism between information structure and tone units, delimited by non-terminal boundaries, i.e. the *information patterning principle* (CRESTI, 2000). Starting from these principles, it becomes possible to carry out corpus-based studies on

spoken language pragmatics based on the perceptual data given by the prosody (CRESTI; MONEGLIA, 2010; MONEGLIA, 2011).

L-AcT assumes, with Austin, that the speech flow is mainly structured in sequences of pragmatically interpretable units, i.e. the Utterances, each one corresponding to the accomplishment of a Speech Act. From the formal point of view, prosody systematically signals the boundaries of each Utterance by means of a conclusive profile; moreover, different illocutions are encoded by different profiles. Therefore, L-AcT provides an explicit criterion for the identification of the fundamental units in the speech flow, based on the retrieval of perceptually relevant prosodic breaks: if an expression is so intonated that it can be pragmatically interpreted in isolation, then it will result in an Utterance.

Nonetheless, the functions of prosody in segmenting the speech flow are not limited to the identification of Utterances and their illocutive values. As a matter of fact, an Utterance can be formed by more than one tone unit, each one signaled by a non-terminal prosodic break. It has been observed that, within the sequence of tone units composing an Utterance, there is usually only one that turns out to be autonomous, while the others can be removed preserving the Utterance interpretability. This prosodic unit corresponds to the Information Unit of Comment, which is therefore necessary and sufficient for the accomplishment of the Speech Act. The expression of the illocutionary value that allows the Utterance interpretation is strictly based on how the Comment unit is prosodically realized,² and does not depend on its morpho-syntactic structure.

L-AcT proposes an original perspective regarding the definition of the information structure of the Utterance, since it is strictly related to the fulfillment of the illocution. Prosodic scanning marks the internal articulation of Utterances, the nucleus of which is constituted by the unit devoted to the accomplishment of the illocution.

To sum up, according to L-AcT prosody plays a crucial role in the realization of the Utterance and in its identification. Prosody is also the way the speaker expresses the illocutionary strength and makes the pragmatic interpretation of Utterances possible.

² The taxonomy proposed by Cresti distinguishes five general illocutionary classes – assertion, direction, expression, rite and refusal – determined by the attitudinal contents of the verbalization (relationship between speaker and interlocutor, emotional content, impulse and representation of action); all participants taking part in the conversation become fundamental objects of the speech act analysis.

2.2 Non-illocutive information units

Information units have either Textual or Dialogic functions. Textual Information Units contribute to the full semantic content of the Utterance. As we already stated, the Comment is the only unit needed to perform the Utterance; the other optional textual units act as a linguistic support for the adequate accomplishment of the Speech Act expressed by the nuclear Informative Unit. Table 1 reports the list of the optional Textual Units, with the tag used in the information labelling and a brief definition.

TABLE 1 – Optional Textual Units

NAME	TAG	BRIEF DEFINITION AND EXAMPLE
Topic	TOP	The domain of application for the speech act accomplished by the Comment. - secondo me / ^{TOP} ne dimostrava di più // ^{COM} <i>[in my opinion / she looked older than her age //]</i>
List of Topics	TPL	A chain of two or more Topics. - gli ordini / ^{TPL} e / ^{SCA} le mansioni / ^{TPL} ti saranno date direttamente da lui // ^{COM} <i>[directives / and / tasks / will be given to you directly by him //]</i>
Appendix of Comment	APC	An integration of the Comment text, either with fillers, repetitions, or delayed information. - era messa male // ^{COM} la nonna // ^{APC} <i>[she was in bad shape / the grandmother //]</i>
Appendix of Topic	APT	An integration of the Topic text. - ma da me / ^{TOP} i' problema / ^{APT} sarà più che altro l' esposizione // ^{COM} <i>[but for me / the problem / will mainly be the exposition //]</i>
Parenthesis	PAR	A meta-linguistic insertion related to the Utterance's content. - se li vedi / ^{TOP} di sicuro / ^{PAR} lo [1] ^{EMP} lo capisci // ^{COM} <i>[if you see them / for sure / you will understand it //]</i>
Locutive Introducer	INT	A specific unit introducing reported speech, a spoken thought, a list, a narration, or an exemplification. - come dire / ^{INT} ci penso io // ^{COM} <i>[you know / I'll take care of this //]</i>

On the contrary, Dialogic Units do not partake in the propositional content of the Utterance and have the function to boost the success of the communicative exchange. They are dedicated, for instance, to keeping the communicative channel open, expressing social cohesion in relation to the interlocutor, and taking or keeping the communicative turn. In Table 2 we list the different Dialogic Units:

TABLE 2 – Dialogic Units

NAME	TAG	BRIEF DEFINITION AND EXAMPLE
Incipit	INP	Opens the communication channel for turn-taking or for performing a contrast. - senti ma / ^{INP} questa è la famosa / ^{SCA} vacanza all' < Elba > ? ^{COM} <i>[listen / is this the famous / holiday on Elba ?]</i>
Conative	CNT	Pushes the addressee to take part in the exchange in an adequate way, inducing him to perform, stop, or avoid a communicative action. - ma che dici / ^{COM} scusami // ^{CNT} <i>[what are you talking about / sorry //]</i>
Phatic	PHA	Ensures that the communication channel stays open and that the dialogical exchange and its reception are maintained. - ecco / ^{PHA} poi questo / ^{TOP} è San Gottardo // ^{COM} <i>[here / then this / is San Gottardo //]</i>
Allocutive	ALL	Identifies the addressee of the Utterance, looking for his attention, and simultaneously establishing a personal connection with him. - queste son belle / ^{COM} mamma // ^{ALL} <i>[these are nice / mum //]</i>
Expressive	EXP	Works as an emphatic support of the exchange, dealing with social cohesion among participants of the communication event. - huf / ^{EXP} fai quello che vuoi // ^{COM} <i>[huf / do what you want]</i>
Discourse Connector	DCT	Connects different parts of the discourse, signaling to the addressee that the discourse is going on. - allora / ^{DCT} all'incirca sei settimane // ^{COM} <i>[so / more or less six weeks]</i>

Empirical studies (see CRESTI 2000; MONEGLIA; RASO, 2014) highlighted the presence of prosodic units that do not bring

any informative value. This is the case of disfluencies or interrupted sequences, as well as “scanning” phenomena. In this latter case, it happens that a single information unit is divided into two or more tone units, mostly for performance reasons; for instance, units with a long textual content may require the performance of two prosodic units. In this case, the prosodic pattern and the information pattern are not strictly isomorphic. The convention adopted in DB-IPIC considers the units on the left as “scanning” units (tag SCA), while the actual information value for the whole unit is annotated only on the last unit. Table 3 reports the list of the tag used for non-informative units.

TABLE 3 – Non-informative Units

NAME	TAG	DEFINITION AND EXAMPLE
Scanning	SCA	A prosodic unit that has no information function on its own, and the content of which is part of a larger IU. - anche qui / ^{TOP} siamo / ^{SCA} a Versailles // ^{COM} <i>[here/ it's / Versailles //]</i>
Interrupted	EMP	An interrupted unit that cannot be evaluated. - e questo è il babbo / ^{COM} quando stavano + ^{EMP} <i>[and this is dad / when they were +]</i>
Time Taking	TMT	A time-taking unit, used for programming needs and/or for keeping the turn. - &he / ^{TMT} no di Virgilio / ^{CMM} della sorella // ^{CMM} <i>[&hem / it's not Virgilio's / it's his sister's //]</i>
Unclassifiable	UNC	An unclassified unit due to insufficient acoustic data. - xxx / ^{UNC} tutto + ^{EMP} <i>[xxx / everything +]</i>

3. Bound Comments and Multiple Comments

As we mentioned earlier, according to L-AcT the Comment unit corresponds to the Utterance nucleus, since it plays the fundamental role of the unit that allows the pragmatic interpretability of the whole sequence.



Usually, a terminated sequence contains only one Comment carrying the illocutionary force of the Utterance. However, it is also possible that more than one independent unit bears an illocutionary value.

This is the case of two different spoken structures, retrieved through a corpus-based analysis.

The first structure is comprised of a chain of units with a homogeneous and weak illocutionary force, i.e. a sequence of Bound Comments (COB). From a prosodic point of view, the characteristic conclusive ending profile of Comments, which brings a singular illocutionary value, is not perceived. In the COB units, the f_0 shape has a continuative profile (which can vary across languages), so that the Comments in the sequence appear, indeed, “bound” together. The illocutionary value is here reduced, since a sequence of Bound Comments is functional to the realization of a unified “story”: the purpose is to build an oral text more than to accomplish a single Speech Act (PANUNZI; SCARANO, 2009). Only the last unit of the chain brings a conclusive prosodic profile, so that it is conventionally signaled as a proper Comment unit (even if it partakes to the whole “bound” sequence).

COBs are typical sequences of monologues and storytelling, in which the exchange between speakers is infrequent. They often coincide with a succession of more than one semantic nucleus held together. Indeed, it is a type of progressive adjunction of speech flow, without a previous and systematic organization of the information. The sequence of Bound Comments allows the formation of another type of basic unit, larger than the Utterance, which has been called Stanza. The main feature of a Stanza is that the sequence of COBs fragments the illocutionary value into various segments which are gradually incremental: they are produced through an adjunctive process, without a strong illocutive activation and prosodic planning. Below are two examples of Stanza taken from the DB-IPIC Italian Minicorpus illustrating the progressive construction of the oral text, both building a narrative sequence. The first (1a) presents a succession of three Comments (two COBs and a COM), and the second shows six units linked together (1b)³:

³ As examples show, other textual and dialogic units can be interposed within a sequence of COBs.


- (1a)  *LIA: la mi' mamma era stata malata /^{COB} era &st [//2]^{EMP} come al solito /^{PAR} era stata all' ospedale /^{COB} e fu proprio il periodo /^{TOP} in cui /^{SCA} mio marito prese /^{SCA} l' azienda /^{SCA} col mi' babbo //^{COM} (ifamcv01_406)
- [*LIA: my mom was sick/ she was/ as usual/ she went to the hospital/ and it was right around the time/ during which/ my husband took over/ the business/ with my dad//]*
- (1b)  *VAL: cioè /^{TMT} niente vabbè /^{PHA} si parte /^{COB} da Firenze /^{COB} eh /^{TMT} si fa i' check-in /^{COB} e si fa direttamente da [1]^{EMP} da Firenze /^{COB} i' check-in /^{COB} eh /^{TMT} per New York //^{COM} (ifammn08_4)
- [*VAL: I mean/ right well/ we fly/ from Florence/ hm/ we check-in/ and directly from Florence/ we check-in/ hm/ to New York//]*


The second structure, Multiple Comments, occurs when a spoken sequence contains two or more Comments, each with its own illocutionary force, held together by a single melodic pattern that connects them. Thus, a higher Information Unit is formed, that is not separable in the interpretation and whose components are unified in a coherent prosodic configuration. It is called Multiple Comment unit (CMM) and it creates an *illocutionary pattern*, i.e. a sequence of illocutive information units within a compositional structure. Each unit has its own characterization and can be, in most cases, pragmatically interpreted.

It is possible to distinguish a CMM from a sequence of independent simple COMs through illocutionary compositional characteristics that are reflected in specific rhythmic and prosodic structures. In fact, this uniform and compositional set of Comments implements special relationships, explained by Cresti (2000) with a classification of meta-illocutionary models that need more than one information units to be executed and produce rhetoric effects, in particular: list, comparison, alternative, and reinforcement relations.


The list pattern is usually a ternary chain (in rare cases binary) of CMMs belonging to the same illocutionary type (e.g. assertions, suggestions, instructions, hypotheses, rhetorical questions, quotations). They contribute to creating a compositional repetition of the same illocutionary force. The main feature of the list is the rhythmic pattern that makes the CMM unitary. Generally, the first segment is prosodically stronger, the second less and the third has a standard conclusive prosodic


profile. The locutive contents of each CMM in the list may vary, but must be semantically coherent. The following are two examples of a list in the form of a Multiple Comment:

 (2a) *ART: pattina /^{CMM} quadrante /^{CMM} fianchi /^{CMM} e maniglia //
CMM(ifamd104_46)
*[*ART: flap/ quadrant/ sides/ and grip//]*



 (2b) *LUI: sul /^{SCA} rispetto /^{CMM} la libertà /^{CMM} quello e quell' altro //
CMM (ipubcv01_420)
*[*LUI: about/ respect/ freedom/ that and that//]*

The comparison pattern is a (usually binary) composition of Comments belonging to the assertive class, or to the total questions. In general, the two locutionary contents are semantically complete, so that the second CMM duplicates the locutionary content of the previous one with some semantic variations, allowing the comparison between the two even in the absence of any explicit lexical mark. Below are two examples of comparison in (3a) and (3b):



 (3a) *CLA: noi la nostra /^{CMM} e loro la loro //^{CMM}(ifammn02_112)
*[*CLA: we have ours/ and they have theirs//]*

 (3b) *SAR: uno per la testata dell' offerta /^{CMM} e l' altra per il corpo
dell' offerta //^{CMM} (ifammn17_11)
*[*SAR: one is for the head of the offer/ and the other for the body
of the offer]*

The alternative pattern is a binary sequence of CMM, largely from the assertive and directive illocutions, which create the composition of two illocutionary forces (e.g. alternative question, alternative instruction, alternative order, total contrast). Normally, both locutive contents are semantically complete, although often the content of the first CMM is filled by a proposition, while the second by a simple phrase or a single word. The content of the two CMMs is always semantically related; see for example (4a) and (4b):

-  (4a) *ALD: perché c'è chi vende /^{SCA} dieci /^{CMM} e chi vende cento ?^{CMM}
(ifammn14, 91)
*[*ALD: why some sell/ ten/ and other sell a hundred?]*
-  (4b) *ASS: bisogna vedere /^{SCA} se lei privilegia una rendita vitalizia /
^{CMM} oppure /^{DCT} un capitale alla scadenza //^{CMM} (ipubdl02_248)
*[*ASS: we must see/ if you prefer an income for life/ or/ a lump sum at the end//]*

Another binary sequence (and the most frequent in production) is the reinforcement pattern, composed by CMMs which belong to a homogeneous illocutionary type; this sequence creates a composition of the two illocutionary forces, which are confirmation, rejection, invitation, agreement, doubt, belief, hypothesis, or related to the class of rites. The locutive content of the first CMM is often filled by an interjection, adverb or stereotyped expression, while the second or the last CMM is filled by a locution that strengthens and makes the message explicit and semantically complete. In other cases, this structure can be inverted, with a first part corresponding to a complete sentence or a phrase and the reinforcement being comprised of a single interjection. There are many cases of reinforcement with functional recall, in which one of the CMM performs the recall function and is combined to a main illocution, usually a directive one. Below, two examples of reinforcement Multiple Comments:

-  (5a) *LIA: già /^{CMM} tu ha' ragione //^{CMM} (ifamcv01_68)
*[*LIA: yes/ you're right//]*
-  (5b) *EST: proprio una chicca /^{CMM} sì //^{CMM} (ifamd115_339)
*[*EST: really doozy/ yes//]*

The two spoken structures just described – Bound Comments and Multiple Comments – characterize together less than the 20% of terminated sequences in spoken Italian (PANUNZI; MITTMAN, 2014).

It is worth highlighting that CMM and COB have different theoretical statuses that reflect on the identification of the reference units for spoken language analysis speech. From a theoretical point of view, the pattern of CMMs composes a sort of higher-level informative unit that

globally functions as a unique Comment; on the contrary, the sequence of COBs forms a chain of independent units that are bound together by adjunction, out of an overall planning. Moreover, as it has been observed by Panunzi and Mittman (2014), the two structures completely differ in their distributional properties (PANUNZI; MITTMAN, 2014). Data from both Italian and Brazilian Portuguese show that COM-Utterances and CMM-Utterances⁴ are similar with regard to their distribution within dialogic interactions and monologic ones, whereas Stanzas (i.e. sequences of COBs) are much more frequent in monologues. The similarities between both types of Utterances (COM and CMM) also extend to their information structure, in which most of the units are simple, i.e. there are no other Information Units except for the Comment (single or Multiple). In contrast, most Stanzas have a complex structure containing at least one optional textual or dialogic IU.

For these reasons, we assume that there is an overall distinction between Utterances (alternatively with COM or CMM as nuclear units) and Stanzas (with COB as nuclear units) as the basic entities for speech segmentation.

4. Examples from DB-IPIC

We investigated the differences between several types of non-terminal breaks, i.e. the ones characterizing Bound and Multiple Comments. As we mentioned above, we carried out an analysis of a set of units extracted from the DB-IPIC Italian Minicorpus. The sample is a qualitative selection composed by 8 Stanzas, with a total of 19 non-terminal COB breaks, as well as 13 Utterances with a total of 18 non-terminal CMM breaks, thus presenting a total of 37 prosodic breaks. The set works as a pilot study for future analysis on a larger collection of COBs and CMMs.

We chose Utterances and Stanzas from 14 different speakers, in conversations (3 speakers), dialogues (6 speakers), and monologues (5 speakers) from the corpus, both familiar (11 speakers) and public (3 speakers). The first criterion for the utterances selection was the audio

⁴ We distinguish Utterance types with respect to the illocutive unit that constitute their nucleus: COM-Utterances are characterized by single Comment nuclear unit, while CMM utterances are characterized by a Multiple Comment nuclear unit.

quality, as we selected the ones with the greatest possible acoustic spectrogram clarity. We then selected speech turns without overlapping.

Selected COBs try to be prototypes of stanzas, with at least three illocutive units⁵ and without final or internal interruptions, since they cannot be confidently evaluated. Whereas, Multiple Comment units were chosen to represent the different CMM types according to Language into Act Theory – list, alternative, comparison, reinforcement. All of the above were patterns of two units, except for four lists of three units.

The following sections will list the transcriptions of analyzed audio tracks, divided into two groups: Section 4.1 contains the Bound Comments and section 4.2 contains the collection of illocutive patterns of Multiple Comments, grouped into the different CMM-types. The beginning of each line gives information concerning the name of the speaker in upper case marked with an asterisk. Then the following transcription of the speech is annotated, with the LABLITA tag set (CRESTI; MONEGLIA, 1997; CRESTI; MONEGLIA, 2005; CRESTI; PANUNZI, 2013), which is a variant of CHAT format for speech transcription (MACWHINNEY, 1991). Following the examples, in brackets, the name of the text to which the segment belongs to in the corpus is specified, with a number used to identify the sequence in the whole text. Each sequence ends with a terminal break and is internally divided into prosodic units through non-terminal breaks. The question mark is used to demarcate a terminated sequence with a rising prosodic profile (as the ones in interrogative or request utterances); double slash, instead, is the standard sign used for terminal breaks, which characterizes conclusive sequences neither interrupted (usually signaled with “+”) nor intentionally suspended by the speaker (MONEGLIA, 2005) (indicated with “...”) Single slash (/) is used for non-terminal breaks. A double or single slash followed by a number, both contained in square brackets,⁶ indicate retracting (i.e. false start, MONEGLIA, 2005) phenomena; *n* corresponds to the number of retracted words. Boundaries of false starts do not contribute to the informational patterning or to the semantic content of the Utterance; hence they are not counted as a proper type of non-terminal breaks.

⁵ The sample contains Stanzas with a maximum of six COBs; however, they are mostly composed by three units.

⁶ In the form of [/*n*] or [//*n*].

4.1 Examples of Bound Comments



- (6) *LIA: la mi' mamma era stata malata /^{COB} era &st [//2]^{EMP} come al solito /^{PAR} era stata all' ospedale /^{COB} e fu proprio il periodo /^{TOP} in cui /^{SCA} mio marito prese /^{SCA} l' azienda /^{SCA} col mi' babbo //^{COM} (ifamcv01_406)

*[*LIA: my mom was sick/ she was/ as usual/ she went to the hospital/ and it was right around the time/ during which/ my husband took over/ the business/ with my dad//]*



- (7) *FRA: e poi /^{INP} perché /^{INT} cioè /^{PHA} non vo' porta' la figliolina lì /^{COB} non la vo' manda' dalla baby-sitter /^{COB} non vo' chiamare i suoceri che son già a i' mare /^{COM} forse //^{PAR} (ifamd112_330)

*[*FRA:and then/ because/ you know/ she doesn't want to bring her kid there/ doesn't want to take her to the sitter/ doesn't want to phone her in-laws who are down the shore/ maybe//]*



- (8) *EST: lei /^{TOP} prima veniva tutte le settimane /^{COB} poi /^{i-COB} purtroppo /^{PAR} gl' è successo un problema alla su' mamma /^{COB} un incidente grosso /^{COB} per cui /^{DCT} ora viene /^{SCA} una volta ogni venti giorni //^{COM}(ifamd115_102)

*[*EST: she/ used to come here every week/ then/ unfortunately/ her mom had a problem/ a serious accident/ so/ now she comes/ once every twenty days//]*



- (9) *CLA: nel quartiere /^{COB} di fratellanza //^{COM}(ifammn02_68)

*[*CLA: in the neighborhood/ between brothers //]*





- (10) *CLA: perché /^{DCT} quella strada la facevano a piedi /^{COB} con la mandria /^{COM} eh //^{PHA}(ifammn03_161)

*[*CLA: because/ that street they were walking/ with the herd/ eh//]*








- (11) *VAL: cioè /^{TMT} niente vabbè /^{PHA} si parte /^{COB} da Firenze /^{COB} eh /^{TMT} si fa i' check-in /^{COB} e si fa direttamente da [1]^{EMP} da Firenze /^{COB} i' check-in /^{COB} eh /^{TMT} per New York //^{COM}(ifammn08_4)

*[*VAL: I mean/ right well/ we fly/ from Florence/ hm/ we check-in/ and directly from Florence/ we check-in/ hem/ to New York//]*



-  (12) *VAL: quindi nulla /^{COB} l' aereo è in orario /^{COB} quindi tranquillamente /^{COB} bene //^{COM}(ifammn08_7)
 [**VAL: and so/ the plane is in time/ and so easy/ ok//*]
-  (13) *GCM: magari con un /^{SCA} testo facile /^{COB}che [1]^{EMP} che ti piace a te /^{COB} e provare /^{i-COM} per esempio /^{PAR} a farli leggere //^{COM}(ipubdl05_188)
 [**GCM: maybe with an / easy book/ that/ that you like/ and try/ for example/ to let them read it//*]

4.2 Examples of Multiple Comments



A) List type:

-  (14) *ART: pattina /^{CMM} quadrante /^{CMM} fianchi /^{CMM} e maniglia //^{CMM}(ifamdl04_46)
 [**ART: flap/ quadrant/ sides/ and grip//*]
-  (15) *NIC: togliamo il resto /^{CMM} ingrandiamo /^{CMM} facciamo solo loro //^{CMM}(ifamd17_279)
 [**NIC: we take the rest off/ we enlarge them/ and do just them//*]
-  (16) *ALD: questo valeva per la Puglia /^{CMM} come pe' la Calabria /^{CMM}o per la Campania //^{CMM}(ifammn14_44)
 [**ALD: that goes for Puglia/ as for Calabria/ or for Campania//*]
-  (17) *SAR: ora niente più lire /^{CMM} niente più dollari //^{CMM}(ifammn17_109)
 [**SAR: now no more lira/ no more dollars//*]
-  (18) *LUI: sul /^{SCA} rispetto /^{CMM} la libertà /^{CMM} quello e quell' altro //^{CMM}(ipubcv01_420)
 [**LUI: about/ respect/ freedom/ that and that//*]





B) Comparison type:

-  (19) *CLA: noi la nostra /^{CMM} e loro la loro //^{CMM} (ifamnn02_112)
 [**CLA: we have ours/ and they have theirs//*]
-  (20) *SAR: uno per la testata dell' offerta /^{CMM} e l' altra per il corpo dell' offerta //^{CMM} (ifamnn17_11)
 [**SAR: one is for the head of the offer/ and the other for the body of the offer //*]

C) Alternative type:

-  (21) *ALD: perché c' è chi vende /^{SCA} dieci /^{CMM} e chi vende cento ?^{CMM} (ifamnn14_91)
 [**ALD: why some sell/ ten/ and other sell a hundred?]*
-  (22) *ASS: bisogna vedere /^{SCA} se lei privilegia una rendita vitalizia /^{CMM} oppure /^{DCT} un capitale alla scadenza //^{CMM} (ipubdl02_248)
 [**ASS: we must see/ if you prefer an income for life/ or/ a lump sum at the end//*]

D) Reinforcement type:

-  (23) *LIA: già /^{CMM} tu ha' ragione //^{CMM} (ifamcv01_68)
 [**LIA: yes/ you're right//*]
-  (24) *ELA: sì /^{CMM} a Roncobilaccio //^{CMM} (ifamcv01_398)
 [**ELA: yes/ in Roncobilaccio//*]
-  (25) *EST: proprio una chicca /^{CMM} sì //^{CMM} (ifamd115_339)
 [**EST: really doozy/ yes//*]
-  (26) *SAR: sì /^{CMM} son io //^{CMM} (ifamnn17_30)
 [**SAR: yes/ it's me//*]

Once the representative sample of Bound Comments and Multiple Comments were selected, we then chose the parameters through which conducting the analysis, as set out below.

5 Acoustic Parameters

In order to approach the issue of differentiating between non-terminal breaks of COB and CMM units, we analyzed phenomena across the prosodic boundaries, both left and right, for all units found after the break.⁷ We took into account different prosodic cues correlating with their perception: f_0 reset; pauses; final lengthening; intensity lowering; initial rush of the following unit (CRUTTENDEN, 1997; HIRST; DI CRISTO, 1998).

F_0 reset was measured in Hertz (Hz). It states differences in pitch range between two adjacent intonation units, namely the difference between the f_0 contours before and after the boundary break (SORIANELLO, 2006). We quote Δf_0 as a percentage of f_0 range in each Utterance/Stanza. Appreciable absolute value of Δf_0 is $>18\%$ (‘T HART, 1981), i.e. at least three semitones. When the f_0 shape changed trend, we annotated the direction of the intonation movement before and after the boundary break: when it was upward, downward or flat on either side of the border. The flat prosodic contour is the case of no significant variation in f_0 values.

Pauses were measured in milliseconds (ms). We evaluated pauses after the boundary break, if present. Appreciable pauses are >180 ms, following Duez (1982, 1985), in which a silent pause is any interval of oscillographic trace where the amplitude is indistinguishable from the background noise – threshold values range from 180 to 250 ms.⁸ According to Moneglia (2005) instead, a perceptively relevant silence in speech continuum has to be longer than 250 ms. Nevertheless, our sample showed no evidence of pauses shorter than 250 ms.

⁷ Our methodological choice was not to distinguish non-terminal breaks on the basis of the next unit since the study is intended as a first step in the formalization of prosodic breaks. Thereafter, the analysis’ aim is to integrate such distinctions, thus taking into account possible prosodic cues determined by characteristics of specific units.

⁸ According to CMU Open Source Speech (<https://cmusphinx.github.io/>) Recognition Software, the smallest pause duration output is 180ms. The same threshold has been adopted by Lundholm Fors (2015).

Final lengthening was measured in milliseconds (ms). It indicates an increase of duration in the last vowel before the boundary break (CHEN, 2007). Appreciable lengthening is >10ms than the mean vowel duration by each speaker. This value was chosen following previous studies (LEHISTE, 1976), where it appears that, in the range of the durations of speech sounds, the just-noticeable differences in duration are between 10 and 40 ms. We used a trimmed mean calculated after discarding the highest and the lowest value, except for the cases of Utterances/Stanzas shorter than 6 V-to-V, where a V-to-V is an acoustic segment delimited by two vowels, measured in seconds from the starting of the first vowel to the starting of the second one (BARBOSA, 2007). The trimmed mean is less sensitive to outliers than the mean, but it still gave a reasonable estimate of central tendency. Where necessary, the outcome was verified also by analyzing other speech segments by the same speaker.

Intensity lowering was measured in decibels (dB). It states a fall in “strength” of articulation. Starting with observations of intensity variation, we recorded the decrease of decibel level just before the boundary breaks (SORIANELLO, 2006).

Initial rush was measured in n(V-to-V)/s. Initial rush indicates a speed up of speech flow at the beginning of a new unit after the boundary break, as a difference of speech rate. The speech rate is useful in order to give the listener a global sense of speed value and to compare various rate levels (OLIVEIRA COSTA; MARTINS-REIS; CÔRREA CELESTE, 2016). We calculated and compared a mean rate per each unit and a local rate for the first two V-to-V segments after the non-terminal break. There is an appreciable acceleration of speech rate – initial rush – with a $\Delta_{\text{Speech Rate}} > 10\%$. The value has been conventionally chosen according to what is detectable to the ear.

The choice of these variables aims at investigating and differentiating COBs and CMMs internal breaks from an acoustic point of view and underlining possible connections between different prosodic cues. In order to take into account the abovementioned parameters, all audio tracks were analyzed through the Praat software and its features of spectrum, pitch and intensity analysis, and the annotation text to sound. They were first divided into Information Units in order to analyze the

f_0 shape; they were then examined with Praat tools as spectrum, pitch, intensity and annotation text-grid tool; the audio tracks were then manually segmented in V toV units.

6 Analysis

This section contains the analysis derived from the study of the parameters described above. The following tables report a synthesis of the results of the analysis per each break. Every mentioned parameter is here mentioned per each examined non-terminal prosodic break. Table 4 shows COBs breaks, while Table (5) is for CMM's breaks.

Guidelines for reading the tables: every break is indicated with the name of the text from which it comes and the ID number, followed by the example number in the above-listed transcriptions (in brackets). Values under the minimum threshold – according to the parameters described above– are in brackets; a blank cell means that the phenomenon does not occur in that specific break. Right and left f_0 trends respect to the boundary are expressed dividing the two paths with a slash. F_0 reset is mentioned as a Δf_0 percentage: a negative Δf_0 percentage is indicated when the reset is up/down and vice versa a positive value for the down/up reset. The “IL” column reports data on intensity lowering; the table takes into account if it is present or not coinciding with breaks. In the same way of f_0 reset, the initial rush is also mentioned as a percentage. As per the vowel duration, in the final lengthening column the number indicated reflects the increase in milliseconds of the last vowel in respect to the mean duration per each speaker.⁹

⁹ The duration measurements shown in the tables have been obtained without a previous normalization of segments. However, we consider the results to be accurate since we did not compare absolute values measurements of different speakers, but only percentages.

TABLE 4 – Analysis of Bound Comment non-terminal breaks

TEXT	n. break	PARAMETERS					
		f_0 trend	f_0 reset	IL	pause after break	rush after break	final lengthening
ifamcv01_406 (6)	6a	down/down		✓			
	6b	flat/flat	(11,4%)	✓	359 ms		+108 ms
ifamd112_330 (7)	7a	down/down		✓		27,7%	+123 ms
	7b	down/down		✓			+53 ms
ifamd115_102 (8)	8a	flat/flat	(6%)	✓			
	8b	flat/flat		✓		15,9%	
	8c	flat/flat		✓			
ifammn02_68 (9)	9a	flat/flat		✓		77,8%	+245 ms
ifammn03_161 (10)	10a	down/up	(-4,8%)	✓		(2,9%)	
ifammn08_4 (11)	11a	down/up		✓		121,2%	+67 ms
	11b	down/flat		✓			+83 ms
	11c	flat/up	-43,3%	✓	351 ms		+134 ms
	11d	flat/flat	(-5,7%)	✓		(3,5%)	
	11e	flat/flat	(-11,4%)	✓			+226 ms
ifammn08_7 (12)	12a	down/flat		✓			
	12b	down/flat		✓			
	12c	down/up		✓			
ipubl05_188 (13)	13a	down/flat		✓			
	13b	flat/flat		✓		55,5%	+77 ms

TABLE 5 – Analysis of Multiple Comment non-terminal breaks, divided in type-groups

TEXT	n. break	PARAMETERS					
		f0 trend	f0 reset	IL	pause after break	rush after break	final lengthening
List							
ifamd104_46 (14)	14a	down/flat	(-3,4%)	✓	371 ms		
	14b	flat/up	(13,1%)	✗		16,3%	
	14c	down/down	(14,3%)	✗		48,7%	
ifamd117_279 (15)	15a	flat/flat		✗		14,5%	
	15b	down/down	28,8%	✗			
ifammn14_44 (16)	16a	down/flat	-37,1%	✗		153,3%	
	16b	down/down	-21%	✓	473 ms		+71 ms
	17a	down/up		✓		11,5%	+54 ms
ipubcv01_420 (18)	17a	down/flat		✓		65,3%	
	17b	flat/flat	(-13,2%)	✓			
Comparison							
	19a	down/down		✓		(1,8%)	
ifammn17_11 (20)	20a	flat/flat	-29%	✓	417 ms		+109 ms
Alternative							
ifammn14_91 (21)	21a	down/down	(13,7%)	✓		26,9%	
ipubdl02_248 (22)	22a	down/up		✓	337 ms		+142 ms
Reinforce							
ifamcv01_68 (23)	23a	flat/flat	26%	✓		75,6%	+50 ms
ifamcv01_398 (24)	24a	down/down		✓		54,1%	
ifamd115_339 (25)	25a	flat/down	-28,2%	✓		69,8%	
ifammn17_30 (26)	26a	up/down		✓			

The analysis shows that COBs have a homogeneous trend to a flat f_0 shape: no big changes of value were recorded in Bound Comments, with just one break with an appreciable f_0 reset (absolute value of Δf_0 around 43%; see example 11a). The sample presents eight cases of flat shape on both sides of the prosodic break (see examples 6b, 8a, 8b, 8c, 9a, 11d, 11e, 13b) and other four with a downward left profile and flat right profile relative to the break (see examples 11b, 12a, 12b, 13a). Furthermore, COBs ending profile was flat in nine of the examples and downward in the other ten, while the start of the new unit assumed more variable shapes, including the upward profile (see examples 10a, 11a, 11c, 12c), without clear preferences.

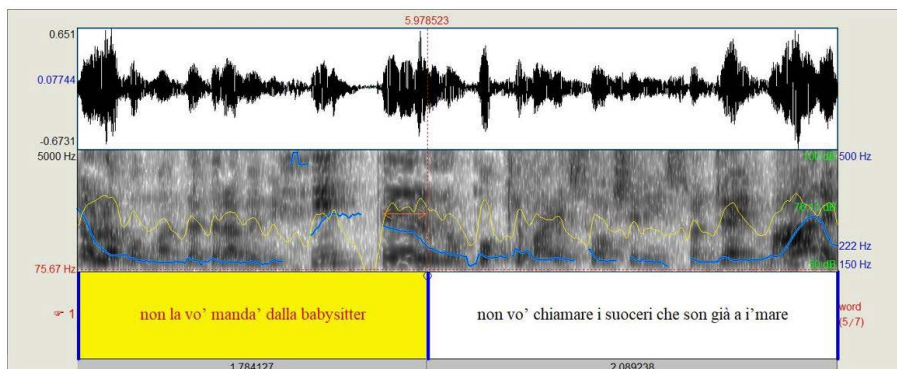
Regarding pauses, none of them after the analyzed breaks was shorter than 250 ms. Two pauses were found after a COB break (see examples 6b, 11c). A lengthening of the final vowel foreran all of these pauses, which were never followed by an initial rush. There were seven other cases of vowel final lengthening. They appeared to be longer than the CMMs final lengthening, to the extent of doubling the duration of the vowel or more in three cases (see examples 6b, 7a, 11e). It is interesting to analyze this data taking into account that the mean vowel duration in COBs was 12 ms longer than in the CMMs unit: the difference was just over the minimum noticeable mark (110 ms for COBs and 98 ms for CMMs). Furthermore, a final lengthening in vowels was present in half of the selection of COB's non-terminal breaks.

Each COB-break had a corresponding intensity lowering, while an initial rush in the following unit followed seven breaks. The increase of speech rate ranged from 3% to 121% values. Following the parameters description above, we consider relevant values when higher than 10%, as the ones observed after five breaks (see examples 7a, 8b, 9a, 11a, 13b), with just one acceleration exceeding 100% (see example 11a).

Mean values of speech rate were not so different between COBs and CMMs and ranged between 5.3-5.5 V-to-V/s.

As an example of COB analysis, Fig. 1 represents the spectrogram of break 7b: the f_0 profile is presented in blue, down before and after the break. The intensity line is shown in yellow, with a decreasing profile before the break. The orange arrow underlines the segment occupied by the last vowel of the first COB, lengthened in comparison with the medium vowel duration of the speaker.

FIGURE 1 – Analysis of COB break 7b



On the other hand, the analysis of the CMM boundaries shows big differences in f_0 ranges between the prosodic breaks. We recorded six appreciable resets, both up/down (see examples 16a, 16b, 20a, 25a) and down/up (see examples 15b, 21a). Their absolute values of Δf_0 varied between around 21-37%, and five $\Delta f_0 < 18\%$, when looking at the absolute values (see examples 14a, 14b, 14c, 18b, 21a); there was a wider f_0 shape variation trend when compared to the COBs group, with the presence of an upward trend before the break too (see example 26a).

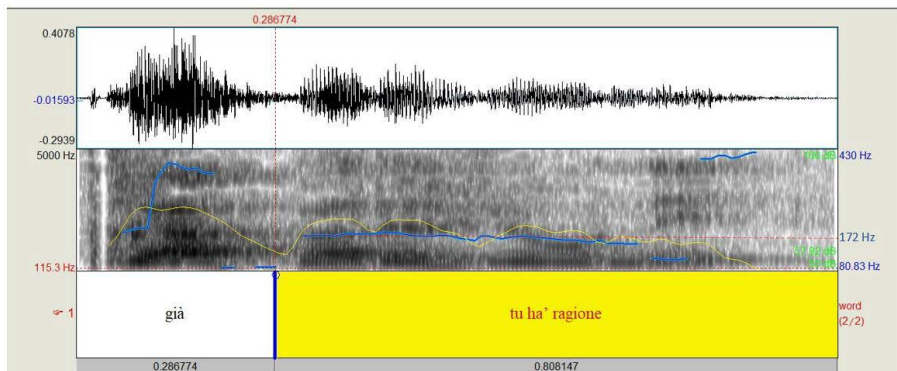
The analyses of the selected audio tracks showed pauses which were noticed after four CMM's breaks (see examples 14a, 16b, 20a, 22a); three pauses were longer than the two recorded between COBs. Only one of them – part of a list-type pattern – was not preceded by a final vowel lengthening (see example 14a). As for COBs, there was initial rush recorded following pauses. Two additional cases of final lengthening occurred before breaks (see examples 17a, 23a). Thus, there were five final lengthening examples in total, with an increase of duration lower than the lengthening observed in COBs, as written above, with two cases of vowel twice the duration of the mean value (see examples 20a, 22a).

When looking at the intensity lowering, five of the analyzed breaks did not present a corresponding decrease in intensity value (see examples 14b, 14c, 15a, 15b, 16a). In all such cases, prosodic breaks were part of a list-type CMM. Moreover, CMMs gave rise to an initial rush in a much easier manner compared to COB's – eleven rushes with a total amount of 18 breaks – with an increasing of speech rate that ranged

between around 2-153% values. Values >10% were observed after ten breaks (see examples 14b, 14c, 15a, 16a, 17a, 18a, 21a, 23a, 24a, 25a), i.e. twice the numbers of COBs, with one acceleration peak >100% (see example 16a).

As an example of CMM analysis, Fig. 2 represents the spectrogram of break 23a: in blue it is possible to see the f_0 profile, which was flat before and after the break. On the right side of the figure, measures of f_0 before and after the break show the f_0 reset from 80.83 Hz – the final point of the first segment – to 172 Hz – the starting point of the second segment. The intensity line is shown in yellow, with a decreasing profile before the break.

FIGURE 2 – Analysis of CMM break 23a



7. Final remarks

Approaching this analysis, we had to face the need for formal parameters to study prosodic features and, especially, the need for fixed thresholds per parameter. It was, therefore, important to specify our set of analytical tools and parameters, based on previous studies but not only. In particular, we chose a conventional value for an appreciable acceleration of speech rate – the initial rush according to what was detectable to the ear, with the aim of better defining the threshold on a perceptual basis.

The aim of this study was to compare features at both sides of the prosodic boundary which are perceived in the speech flow and, in view of the results, the analysis suggests some correlations between the different parameters.

To sum up, every pause after a non-terminal break, whether a COB's or a CMM's ones, is always preceded by a final lengthening of the last vowel of the relevant unit and never followed by an increase in speech rate. Furthermore, the coincidence between intensity lowering and a non-terminal break fails only for list-type Multiple Comments and it is easier to observe a final lengthening at the end of a Bound Comment unit.

As we explained, one of the main characteristics of the Bound Comment is that the end of the f_0 shape continues in the following units, so that the Comments appear, namely, bound together. Thus, in line with our expectation, the distinctive features of Bound Comment are non-terminal breaks with a flat trend of f_0 shape before the boundary, with a low number of f_0 reset, while, on the other hand, Multiple Comments vary between different f_0 shapes on either side of the boundary, which are rarely flat and most of them have a reset.

Furthermore, vowel lengthening and a no rushing speech rate both have an effect in perceiving the prolongation of one COB into another: the results indicate therefore that initial rush is a typical feature of Multiple Comments, while the lengthening of the last vowel of the unit is easier to find at the end of a Bound Comment compared to CMMs.

Moreover, the decision to divide the results about CMMs in different types has been helpful in order to underline the differences between patterns, such as the contrast between lists and the other types concerning the co-presence of non-terminal break and intensity lowering. Of course, it is necessary to replicate the tendencies which were found in our sample by investigating a larger set of consistent cases.

Since our analysis was carried out on a pilot sample, it is clear that these hypotheses need to be tested on a larger set of spoken sequences. It will be interesting to analyze whether or not new observations will reflect the partial results of this sample, in particular concerning the differences in values of initial rush between COBs and CMMs and the properties change between CMM-patterns. Further examples could confirm the COB correlation with the absence of an upward f_0 profile just before the prosodic break, as was suggested in our sample.

The absence of a rising profile is a remarkable result, given that the typical signal of continuity between prosodic units requires an upward direction on the last section of the f_0 profile. Instead, our sample shows that the last syllable does not present a rising phenomenon, but rather the profile is downward or flat. Future studies could deepen the

observation of the previous syllables, the tonic one in particular, as well as the comparison with non-COB continuity signal or rising profile.

Thus, our aim is to extend the analysis to the entire DB-IPIC Italian Minicorpus. This work, implemented with an automatic segmentation of spoken tracks in V-to-V, could also lead to an improved identification and the tagging of Bound Comments and Multiple Comments in DB-IPIC, also when interrupted sequences occur in the speech flow.

Authors' contribution

The authors conceived and discussed together all the content of this paper. However, their own contribution can be specified as follows: Alessandro Panunzi directed the research, provided the examples from DB-IPIC corpus, and wrote Sections 1-3; Valentina Saccone carried out the prosodic analysis and wrote Sections 4-7.

References

AUSTIN, J. L. *How to Do Things with Words*. Oxford: Oxford University Press, 1962.

BARBOSA, P. A. Análise e modelamento dinâmicos da prosódia do português brasileiro. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 15, n. 2, p. 75-96, 2007. Available from: <<http://www.periodicos.letras.ufmg.br/index.php/relin/article/view/2449/2403>>. Access on: Jan. 10, 2018.

BOERSMA, P.; WEENINK, D. *Praat: Doing phonetics by computer*. 2005. Software. Available from: <<http://www.praat.org/>>. Access on: Jan. 10, 2018.

CHEN, A. Language specificity in the perception of continuation intonation. In: GUSSENHOVEN, C.; RIAD, T. (Org.). *Tones and Tunes*. v. II: Experimental Studies in Word and Sentence Prosody. Berlin: Mouton De Gruyter, 2007. p. 107-141.

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

CRESTI, E.; MONEGLIA, M. (Org.). *C-ORAL-ROM. Integrated reference corpora for spoken romance languages*. Amsterdam: John Benjamins Publishing Company, 2005.

CRESTI, E.; MONEGLIA, M. Informational patterning theory and the corpus based description of Spoken language. The compositionality issue in the Topic Comment pattern. In: MONEGLIA M.; PANUNZI A. (Org.). *Bootstrapping Information from Corpora in a Cross Linguistic Perspective*. Firenze: Firenze University Press, 2010. p. 13-46.

CRESTI, E.; MONEGLIA, M. L'intonazione e i criteri di trascrizione del parlato. In: BORTOLINI, U.; PIZZUTO, E. (Org.). *Il progetto CHILDES Italia*. Pisa: Il Cerro, 1997. v. II, p. 57-90.

CRESTI, E.; PANUNZI, A. *Introduzione ai corpora dell'italiano*. Bologna: Il Mulino, 2013.

CRUTTENDEN, A. *Intonation*. 2. ed. Cambridge: Cambridge University Press, 1997.

DUEZ, D. Silent and Non-Silent Pauses in Three Speech Styles. *Language and Speech*, Kansas, v. 25, n. 1, p. 11-28, 1982.

DUEZ, D. Perception of silent Pauses in Continuous Speech. *Language and Speech*, Kansas, v. 28, n. 4, p. 377-389, 1985.

HIRST, D.; DI CRISTO, A. (Org.). *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press, 1998.

LEHISTE, I. Suprasegmental Features of Speech Use. In: LASS, N. J. (Org.). *Contemporary Issues in Experimental Phonetics*. New York: Academic Press, 1976. p. 225-239.

LUNDHOLM FORS, K. *Production and Perception of Pauses in Speech*. 2015. 141p. Dissertation (Doctoral) – University of Gothenburg, 2015. Available from: <https://gupea.ub.gu.se/bitstream/2077/39346/1/gupea_2077_39346_1.pdf>. Access on: Jan. 10, 2018.

MACWHINNEY, B. *The CHILDES Project*. Tools for analysing talk. Hillsdale, NJ: Lawrence Erlbaum Associates, 1991.

MONEGLIA, M. The C-ORAL-ROM Resource. In: CRESTI, E.; MONEGLIA, M. (Org.). *C-ORAL-ROM. Integrated reference corpora for spoken romance languages*. Amsterdam: John Benjamins Publishing Company, 2005. p. 1-70.

MONEGLIA, M. Spoken corpora and pragmatics. *Revista Brasileira de Linguística Aplicada*, Belo Horizonte, v. 11, n. 2, p. 479-519, 2011.

MONEGLIA, M.; RASO, T. Notes on the Language into Act Theory. In: RASO T.; MELLO H. (Org.). *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins Publishing Company, 2014. p. 468–495.

NICOLAS MARTINEZ, C. *Cor-DiAL*. Corpus oral didáctico anotado lingüísticamente. Madrid: Liceus, 2012.

OLIVEIRA COSTA, L. M.; MARTINS-REIS, V. de O.; CÔRREA CELESTE, L. Methods of analysis speech rate: a pilot study. *CoDAS*, São Paulo, v. 28, n. 1, 2016. Available from: <http://www.scielo.br/pdf/codas/v28n1/en_2317-1782-codas-28-01-00041.pdf>. Access on: Jan. 10, 2018.

PANUNZI, A.; GREGORI, L. DB-IPIC. An XML database for the representation of information structure in spoken language. In: PANUNZI A.; RASO T.; MELLO H. (Org.). *Pragmatics and Prosody*. Illocution, Modality, Attitude, Information Patterning and Speech Annotation. Firenze: Firenze University Press, 2012. p. 133-150. Available from: <<http://www.lablita.it/app/dbipic/>>. Access on: Jan. 10, 2018.

PANUNZI, A.; MITTMAN, M. The IPIC resource and a cross-linguistic analysis of information structure in Italian and Brazilian Portuguese. In: RASO T.; MELLO H. (Org.). *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins Publishing Company, 2014. p. 129-151.

PANUNZI, A.; SCARANO, A. Parlato spontaneo e testo: Analisi del racconto di vita. In: AMENTA L.; PATERNOSTRO G. (Org.). *I parlanti e le loro storie*. Competenze linguistiche, strategie comunicative, livelli di analisi: Atti Del Convegno Carini-Valderice. Palermo: Centro di studi filologici e linguistici siciliani, 2009. p. 121-132.

RASO, T.; MELLO, H. (Org.). *C-ORAL-BRASIL I*: Corpus de referência do Português Brasileiro falado informal. Belo Horizonte: Editora UFMG, 2012.

SORIANELLO, P. *Prosodia*. Modelli e ricerca empirica. Roma: Carocci, 2006.

‘T HART, J. Differential sensitivity to pitch distance, particularly in speech. *The Journal of the Acoustical Society of America*, USA, v. 693, p. 811-821, 1981.



Syntax, Prosody, Discourse and Information Structure: The Case for Unipartite Clauses. A View from Spoken Israeli Hebrew

*Sintaxe, prosódia, discurso e estrutura informacional:
o caso das orações unipartidas.
Uma visão do hebraico falado em Israel*

Shlomo Izre'el

Tel Aviv University, Tel Aviv, Israel

izreel@tauex.tau.ac.il

Abstract: The canonical view of clause requires that it include predication. Utterances that do not fit into this view because they lack a subject are usually regarded as elliptical or as non-sentential utterances. Adopting an integrative approach to the analysis of spoken language that includes syntax, prosody, discourse structure, and information structure, it is suggested that the only necessary and sufficient component constituting a clause is a predicate domain, carrying the informational load of the clause within the discourse context, including a “new” element in the discourse, carrying modality, and focused. Utterances that have not been hitherto analyzed as consisting of full clauses or sentences will be reevaluated. The utterance, being a discourse unit defined by prosodic boundaries, can thus be viewed as the default domain of a clause or a sentence, when the latter are determined according to the suggested integrative approach.

Keywords: syntax; clause structure; information structure; discourse; context; prosody; utterance; history of linguistics; spoken Israeli Hebrew.

Resumo: A posição canônica sobre as orações requer que elas contenham uma predição. Enunciados que não se encaixem nessa visão porque não possuem um sujeito são usualmente considerados elípticos ou como enunciados não-oracionais. Adotando uma visão integrativa para a análise da língua falada, que inclui a sintaxe, a prosódia, a estrutura discursiva e a estrutura informacional, sugere-se que o único componente

constituente necessário e suficiente para uma oração é um domínio predicativo, o qual carregue a carga informacional da oração no contexto do discurso, incluindo-se um “novo” elemento no discurso, que carregue modalidade e foco. Enunciados que até então foram classificados como não sendo orações ou sentenças completas serão reavaliados. O enunciado, sendo uma unidade discursiva definida por fronteiras prosódicas, pode assim ser visto como o domínio de uma oração ou sentença por excelência, quando estas são determinadas através da abordagem integrativa sugerida.

Palavras-chave: sintaxe; estrutura oracional; estrutura informacional; discurso; contexto; prosódia; enunciado; história da Linguística; hebraico israelense falado.

Submitted on January 11th, 2018.

Accepted on May 12th, 2018.

1 “It’s all Greek to me”

Linguistics ... has an analytical system based on categories that were established at the beginnings of its history, between 400 BC and 600 CE. This system has been transposed into the common epistemological system, the collective knowledge, in almost all European cultures. ... The grammatical activities leaned on the only language considered as such, namely Greek, and, when needed, also on Latin. ... The history of Linguistics since the beginning of the 16th century might well be written as a history of rejection and repression of all linguistic phenomena that are not in accordance with the system of presuppositions of European linguistics. (EHLICH, 2005, p. 104-106; my translation)

The dawn of linguistic research had its roots in philosophical, ontological and logical traditions of ancient Greece, notably those founded by Plato and Aristotle (5th-4th centuries BC). In Plato’s *Sophist*, one finds the following discussion of what we can now refer to as ‘sentence’ or ‘clause’:¹

¹ The Greek term *λόγος*, which relates to the semantic field of speech (LSJ s.v.), may represent a broad range of speech units, and has been translated below as either ‘discourse’ or ‘sentence’, depending on the context. For the term *ῥήμα*, most commonly translated and conceived as ‘verb’, see below.

[D]iscourse is never composed of nouns alone spoken in succession, nor of verbs spoken without nouns.

...

[F]or in neither case do the words uttered indicate action or inaction or existence of anything that exists or does not exist, until the verbs are mingled with the nouns; then the words fit, and their first combination is a sentence, about the first and shortest form of discourse.

...

A sentence, if it is to be a sentence, must have a subject; without a subject it is impossible.

...

And if there is no subject, it would not be a sentence at all; for we showed that a sentence without a subject is impossible. (PLATO, *Sophist*, §§262a-263d; translation by FOWLER, 1921)

In a similar vein, Aristotle, Plato's disciple, defined *ῥῆμα* as "a sign of what is being said on another thing". Aristotle further requires that *ῥῆμα* consignify time. (ARISTOTLE, *Περὶ Ἑρμηνείας*, 16^b6; ARENS, 1984, p. 22, §17).

The Greek word *ῥῆμα*, which originally means anything spoken, has most commonly been interpreted and translated as 'verb', being an anachronistic interpretation of *ῥῆμα* as a technical term. This interpretation probably originated in Aristotle's further requirement from *ῥῆμα* to consignify time.

It should be noticed at this juncture that Aristotle, whose impact on the development of Western linguistics cannot be underestimated (ARENS, 1984, p. XX; ALLAN, 2004), used only a well distinguished and accommodated application of language for his needs, i.e., ontology and logic (ILDEFONSE, 1994). Aristotle explicitly states that

not every sentence is a statement-making sentence, but only those in which there is truth or falsity. There is not truth or falsity in all sentences: a prayer is a sentence but is neither true or false. The present investigation deals with the statement-making sentence; the others we can dismiss, since consideration of them belongs rather to the study of rhetoric or poetry. (ARISTOTLE, *Περὶ Ἑρμηνείας*, 16^b33; translation by ACKRILL, 1961, p. 45-6)

Thus, the language of logic is different in goals from ordinary language and it may well differ in form, e.g., in the requirement that *ῥῆμα*

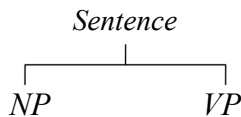
as a logical predicate consignify time, noticing that in Ancient Greek a predicate does not necessarily have to be a verb. Still, Western linguistics has transmitted the original Greek term *ῥῆμα*, via its Latin translation *verbum*, which, like Greek *ῥῆμα*, originally meant ‘anything spoken’, to become the technical term as we understand it today. This point may add further support to the claim that logic rather than language was the root upon which linguistic thinking has had its beginnings. This need not concern us at the moment, although this conception of the term has influenced Western syntactical thinking to the point that any sentence (or clause) is believed to require the presence of a verb, which is not the case in a plethora of languages around the world, including European ones. We shall return to this issue later.

At this point, our interest lies with the requirement to have at least two components in a simple sentence or clause: a subject and a predicate. As mentioned, this requirement has its bases in ancient philosophy and logic, which was carried on to be a basic requirement in the Western study of syntax ever since (SANDMANN, 1979, especially Part II; SEUREN, 1998, §§2.6.3; p. 512).

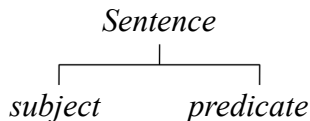
Indeed, *sentence* (or *clause*) is defined even today in terms of subject and predicate, literally as consisting of N(oun)P(hrase) and V(erb)P(hrase), very much like in the dawn of linguistics and its forerunners in philosophy. The most notorious conception of sentence structure was formulated as

$$\text{Sentence} \rightarrow \text{NP} + \text{VP}$$

or



(CHOMSKY, 1957, p. 26-27) or, using syntactically “functional notions”, as



(CHOMSKY, 1965, p. 68-69). This formulation (sometimes allowing some modifications) has become the basis for analyses of sentence structure to date (VAN VALIN; LAPOLLA, 1997, ch. 2; CULICOVER;

JACKENDOFF, 2005, p. 99; GENETTI 2014, p. 121; among many others). Subject and predicate are thus regarded as the very core components of clause structure.

This requirement has been faced time and again with linguistic reality. Hence, recent definitions of clause may make some concessions; (e.g.): “A clause can be defined as a syntactic unit *typically* consisting of a verb (...), its noun arguments, and optional adverbial elements (...)” (GENETTI, 2014, p. 130, my emphasis; note that only adverbial elements are said to be optional).

Interestingly enough, already during the philosophical era and before the rise of grammatical tradition, the Stoics distinguished between complete and incomplete (*ελλιπή*) λεκτά ‘sayables’, among the latter were predicates without a specified subject (LONG; SEDLEY, 1987, v. I, p. 199-200; cf. BLANK; ATHERTON, 2009, p. 315):

Sayables, the Stoics say, are divided into complete and incomplete, the latter being ones whose linguistic expression is unfinished, e.g. ‘[Someone] writes’, for we ask ‘Who?’ In complete sayables the linguistic expression is finished, e.g. ‘Socrates writes.’ So incomplete sayables include predicates, whereas ones that are complete include propositions, syllogisms, questions and enquiries. (DIOGENES LAERTIUS, 7.63 *apud* LONG; SEDLEY, 1987, v. I, p. 196; Greek original: *op. cit.*, v. II, p. 199).

The concept of *ellipsis*, having arisen within prescriptive orientations of language studies, is already found in the first study of syntax by Apollonius Dyscolus (2nd century CE). For Apollonius, “the deleted words have a virtual presence, which will be revealed by the requirements of the sentence” (APOLLONIUS DYSCOLUS, *Περὶ συντάξεως*, §42; cf. HOUSEHOLDER, 1981, p. 33; LALLOT, 1997, p. 108-109). As of today, the common practice has remained very much the same as the one adopted by Apollonius: a sentence (or clause) must have a predesigned form with required components. If these are not found in an actual sentence, the sequence is being regarded as an elliptical sentence, as if a virtual component is represented in the sequence as a zero component; alternatively, it will be regarded as a non-sentential utterance (BENAYOUN, 2003; STANTON, 2004; CULICOVER; JACKENDOFF, 2005; FOLEY, 2006; WINCKLER, 2006; REICH, 2011; GINZBURG, 2012; MERCHANT, 2015; among many others).

As regards our case here, genuine linguistic observation tells us that subjectless clauses are amply attested in spoken languages (e.g., GIVÓN, 1983; BIBER *et al.*, 1999, §§14.3.3-4; CRESTI, 2005; cf. IZRE'EL, 2005, p. 4-5). Thus, according to the common view, the study of spoken languages allow for many non-sentential utterances. Indeed, Carter & McCarthy (2006, p. 490) explicitly claim that “[t]he sentence is a unit of grammar, and must be grammatically complete (i.e. it must have at least one main clause). The utterance is a unit of communication. It [...] does not need to be grammatically complete”. Biber *et al.* (1999, ch. 14) use the term “non-clausal” for units that do not conform to the traditional definition of a clause, yet nevertheless feel the need to coin an “umbrella term ‘C-units’ for both clausal and non-clausal units; i.e., for syntactically independent pieces of speech” (p. 1070). This was done precisely because many of the units used in everyday speech do not fit in the “received receptacles”, to use Sinclair’s metaphor in his review of this magnum opus (SINCLAIR, 2001, p. 357; see note 3 below).

Givón, in his book *The Story of Zero*, comments as follows:

When coded as a **verbal clause** in actual communication, the mental proposition may only weakly resemble the full fledged Aristotelian proposition or its Chomskian deep-structure equivalent, with obligatory subject and verb and optional objects and adverbs. In spontaneous spoken language, the mental proposition often appears as an elliptic, truncated structure, with zeroed out arguments or even a zeroed out verb. (GIVÓN, 2017, p. 28-29)

Lee *et al.* (2009), drawing attention to the fact that in many languages the lack of subject in spoken discourse is pervasive, find the same tendency in English, concluding that

[s]uch phenomena in conversation are not syntactic anomalies... Unfortunately, linguists have neglected this sort of grammar and language or have imposed inappropriate categories from writing. ... We must conclude, then, that the “omission” of subjects (and other arguments) is not an omission at all but a natural and ordinary practice in English grammar that has simply been overlooked because of our reliance on artificially manipulated grammar. If anything, overt subjects are “additions” to English grammar. (LEE *et al.*, 2009, p. 106)

This bias towards written language analysis is reflected in the title of a book by Per Linell, *The Written Language Bias in Linguistics*. Linell claims, *inter alia*, that “‘elliptical’ sentences are fully functional and sufficient for their communicative purposes, given the relevant sequential positions and activity contexts in which they occur” (LINELL, 2005, #41, p. 74).

For Givón, it is rather the bias towards competence:

Lastly, a more general – theoretical or methodological – lesson to be drawn from this study concerns the linguist’s bias toward ‘competence’ data. Reflective, well edited, written English may well be an empirical fiction. As much as we love it as writers, as linguists we may have to stop basing our theories of natural language on this quaint artifact. (GIVÓN, 2017, p. 156)

Some languages are notoriously sparing in the use of subjects (see, *inter alia*, KIBRIK, 2011, §3.4; GIVÓN, 2017, ch. 5). This is especially prominent in languages of Asia and the Pacific (GIVÓN, 2017, p. 130). For example, a textbook of Japanese for foreign students states as follows: “Clauses without subjects are very common in Japanese; Japanese speakers actually tend to *omit* subjects whenever they think it is clear to the listener what or who they are referring to” (BANNO; OHNO; SAKANE; SHINAGAWA, 1999, p. 14; my emphasis).

Japanese linguists tend to refer to clauses without subjects as if the subjects are “missing”, terming subjectless clauses as showing “nominal ellipsis” or as consisting of “null anaphora” or “zero pronouns” (TSUJIMURA, 2007, p. 255-256; IWASAKI, 2013, p. 279). Iwasaki states that

[s]ince Japanese does not have any co-referencing system between arguments and the predicate, the process of zero anaphora is largely pragmatic, and contextually retrievable information can be, more often than not, unexpressed. (IWASAKI, 2013, p. 279, referring to OKAMOTO, 1985)

Iwasaki comments on the use of terms:

Although the terms such as ellipsis and zero anaphora are used in this chapter, it is more accurate to state that not expressing a noun argument is an unmarked case, both in spoken and written discourse, and only pragmatic necessity such as disambiguation and initial mention requires an overt noun in discourse. (IWASAKI, 2013, p. 279, note 3)

Indeed, Japanese linguistics has drawn much from Western traditions (CHUNG, 2013, §10.4), to the extent of using structural trees as in Generative linguistics and other schools in contemporary linguistics (e.g., TSUJIMURA, 2007, p. 255-256).

A notable tradition struggling with analyses of unipartite clauses, i.e. clauses consisting of only a predicate domain, has originated in Francophone scholarship. According to this tradition, utterances that do not fit the concept of predication between two components are still considered sentences, where subjects are not required at all; rather, predicates and modality form complete sentences (cf., e.g., BALLY, 1965, §§49, p. 61-65; TESNIÈRE, 1966 [English: 2015], chs. 45-46, 73, 75, 77; LE GOFFIC, 1993, §351; LEFEUVRE, 1999, *Troisième partie*; BLANCHE-BENVENISTE, 2006, §3). See further §5 below.

As for Israeli Hebrew, the language I am using here as a test-case, it should be mentioned that although spoken Hebrew does not dispense with subjects at the rate Japanese and other subject-sparing languages do, still unipartite clauses are quite frequent in spontaneous spoken Hebrew. For an illustration of the find, I chose a 20'12" conversation consisting mostly of small narratives (uttered by speaker 1). As Table 1 shows, more than half of the substantive units and more than 90% of the regulatory units do not manifest predication.² Thus, units without predication form the majority in the sample. Such units will be analyzed as unipartite clauses, consisting of only a predicate domain.

² Substantive units are those which carry the contents of the discourse; regulatory units are those which regulate the discourse flow (CHAFE, 1994, p. 63-64).

TABLE 1 – Units with and without predication

	speaker 1			speaker 2		
	total	+predication	-predication	total	+predication	-predication
utterances:	344			186		
incomplete:	-21			-6		
complete:	323	145	178 — 55%	180	80	100 — 56%
substantive:	283	140	143 — 51%	130	76	54 — 41.5%
regulatory:	40	3	37 — 92.5%	50	4	46 — 92%

As we have seen above, the concept of bipartite structure of sentence/clause goes back to ancient Greece, be it to its language or logic (cf. further MAUTHNER, 1901-1902, v. III, p. 4; 1907, p. 96-97; LENK, 1993; GIL, 2012, p. 330). It is Greek philosophy out of which Western linguistics has sprung, starting with the study of the language of Ancient Greece, spreading to the study of Latin, and from there to the study of other European languages and much beyond. As regards linguistics, it indeed seems that it's all Greek to us. The burden of grammatical tradition may be too heavy. Perhaps one should, once and for all, dispense with this burden and start — or rather restart — take a fresh look at language, using authentic linguistic data, as has been sporadically called for along the history of linguistics (see further below, §7).

2 Premises

Before bringing forward my analysis of unipartite clauses in spoken Hebrew, I should state here the premises that serve as a guide for my work on spoken language (IZRE'EL, 2012, §1; IZRE'EL, 2018, §§1,2):

- Language is, first and foremost, a tool of expression and communication. Its most frequent manifestation is human communication.
- Language should be studied for its own sake. A corollary of this demand is that linguistic analysis must detach itself from any dependence on other disciplines, notably logic.

- Spoken language varieties, notably the language of everyday conversation, are the most frequently used among all linguistic systems. It is this capacity of spoken language that lends it the power to have its impact on all other linguistic systems and their development.
- Given the prominence of spoken language in human communication, proper linguistic attention must be drawn to the spoken varieties of language, notably spontaneous ones.
- Spoken language must be analyzed according to its own properties. We must detach ourselves from any preconceptions about the structure of language based on its written forms.
- Corpus-driven approach. viz., building up a theory of language from actual data (TOGNINI-BONELLI, 2001) is to be preferred over corpus-based approach, viz., looking for data to establish a preconceived theory.³
- Corpus data reflect the perceived language rather than the produced one. Therefore, linguistic description and analysis based on corpus data can lean solely on data as heard rather than as generated by the speaker, as we do not have direct access to the linguistic system that had generated the actually produced speech.⁴
- Language is intimately related to discourse, so that it will express only what is needed to be expressed within the discourse context, be it linguistic or extra-linguistic.
- Accordingly, language cannot be disconnected from the discourse for the sake of analysis.

³ “To me a corpus of any size signals a flashing neon sign ‘Think again’, and I find it extremely difficult to fit corpus evidence into received receptacles ... the language obstinately refuses to divide itself into the categories prepared in advance for it” (SINCLAIR, 2001, p. 357). “It is not about using spoken French to illustrate a theory, but finding a theory that allows to approach the data of spoken French” (BLANCHE-BENVENISTE; JEANJEAN, 1987, p. 90; my translation).

⁴ This perspective does not contradict the possibility to look into cognitive processes while scrutinizing the received materials (see, e.g., the remarks by GIVÓN, 1992, especially §6 and §8; also the methodology used by KIBRIK, 2011).

- Notwithstanding its mutual-relationship with elements that are either external to the linguistic system or external to the immediate discourse, language is a system on its own, and must be analyzed accordingly.
- Referents are not part of the linguistic structure; they may or may not be represented in the discourse at any time. Furthermore, potential arguments need not be represented in the syntactic form.
- Taking the point of view of the recipient, there can be no question about ellipsis at all.

Using these premises as guidelines for my work, I will try to determine the notion of *unipartite clause*. Before doing that, let me draw a few guiding lines on the structure of Hebrew that will set up the ground for this undertaking.


3 C'est de l'hébreu pour moi

When French people say *C'est de l'hébreu pour moi* ("It's Hebrew to me"), they mean exactly what Americans mean when they say "It's (all) Greek to me". Having suggested that Greek, in its peculiar way, has blocked our understanding of other languages, or, rather, made us look at other languages and language in general taking the point of view of Greek (language or philosophy) (§1), let us see whether the study of Hebrew can suggest some other ways for the analysis of clause structure.


Hebrew, like many other languages, does not require a verb to be its predicate. In fact, *any* part of speech (save bare prepositions, except for some special cases) can form a predicate: nominal (substantives, adjectives, participles), pronominal (personal pronouns, demonstratives, interrogatives and other pronouns), adverbs and prepositional phrases, as well as larger phrases, clauses and other types of syntactic complexes (IZRE'EL, 2012, §3). Some examples:⁵

⁵ The data for this research is drawn from spontaneous speech recordings collected for *The Corpus of Spoken Israeli Hebrew (CoSIH)* <<http://cosih.com/english/index.html>>. References follow the system used in *CoSIH*; speakers are referred to as sp1, sp2, etc. Excerpts that are not retrievable from *CoSIH*'s website are referred to by record reference followed by time measures (exx. 10, 14, 17).


Substantive:

- (1) ze **ha = sa'lon** ||
 DEM[SGM] DEF=living.room 
 'This is the living room.'
 (C842_sp1_166)


Adjective:

- (2) a'val ze **ja'fe** ||
 but DEM[SGM] **beautiful**[SGM] 
 'But this is beautiful.'
 (C711_1_sp1_024)

Active participle:

- (3) v **ha = 'otobus** **ko'fets** |
 and DEF=bus(SGM) **jumping**[SGM] 
 'and the bus is jumping,'
 (OCh_sp1_176)

Prepositional (adverbial) phrase:


- (4) a'ni **be = 'kurs** ||
 I **in=course** 
 'I am taking a course.'
 (OCD_3_sp1_059)

Transcription is usually broad phonetic, with some attention to the phonological system. Phonological input is added mainly in the representation of /h/, which is omitted in most environments in contemporary spoken Hebrew, and in the representation of some occurrences of /j/, which may also elide in certain environments. For typographic and reading convenience, the rhotic phoneme, which in standard Israeli Hebrew is uvular, is represented as *r*; the mid vowels are represented as *e* and *o*, although their prototypical respective pronunciations are lower. Two successive vowels are separated by a syllabic boundary, e.g., 'bait 'house' is to be read 'ba.it; diphthongs are indicated by vowel+semi-vowel (in both directions), e.g., aj, ja. Glossing follows, mutatis mutandis, the Leipzig Glossing Rules <<http://www.eva.mpg.de/lingua/resources/glossing-rules.php>>.


Notation: | minor boundary; || major boundary; / major boundary with "appeal" tone; — fragmentary (truncated) prosodic unit; - truncated word (cf. IZRE'EL, 2002, following in essence DU BOIS *et al.*, 1992).

Predicates in Exx. 1-7 are indicated by boldface characters.


Existential negation:

- (5) **en** kvif || 
NEG.EXT paved.road
 ‘There are no paved roads.’
 (OCh_sp1_179)


Complex:

- (6) ma =ʃ= tsa'riχ liv'dok ba'sof | **ze ma ro'tsim ʃ=ji'hje** || 
 what=that=necessary to.check in.the.end **this what want.PLM that=it.will.be**
 ‘That is, what has to be checked in the end is what one wants that will take
 place.’
 (OM_sp6_004-005)

One may be surprised that verbs are not listed among the predicate types. The reason is that a verb makes a whole clause in itself, as it comprises both a pronominal subject morpheme and a verbal predicative stem:

- (7) **hits'liχ-a** / 
succeeded-3SGF
 ‘Did she succeed?’
 (C714_sp1_096)

In all the cases above, the cited clauses are bipartite. As mentioned above (§1), spoken Hebrew is ample with utterances without predication. Ex. 8 illustrates this type of utterances, which I am suggesting that they be regarded syntactic units, viz., clauses. Sp1 had told sp2 about a ride he had taken in Mongolia on a local breed of horses, and sp2 suggested that they were mules rather than horses. Sp1 insisted that this kind of animal is a genuine horse, and sp2 now responds by a verifying question:

- (8) sp2: sus ma'maʃ /
 horse real 
 '(Is it) a real horse?'
- sp1: sus sus |
 horse horse
 '(It is) a real horse,'
- rak jo'ter na'muχ ||
 only more short
 'but shorter.'
- rag'laim mekutsa'rot ka'ele ||
 legs shortened sort.of
 '(It has) sort of shortened legs.'
 (OCh_sp2_091; sp1_286-288)

In this exchange, quite typical of Hebrew casual talk, none of the units conforms to the common definitions of *clause* as a unit consisting of both subject and predicate. Taking the point of view of the recipient (§2), I would rather not refer to nonexistent elements as if elided or missing. I will try to find a path through which we can reach a unified theory that will encompass all the evidence provided by spontaneous speech data including units that do not include predication and therefore are usually not regarded as (complete) clauses. In other words, I will try to accommodate unipartite clauses into a unified theory of clause structure (IZRE'EL, 2012).

Taking into account the discussion hitherto, we may bring forth the following questions:

If a predicate does not have to be a verb, so that arguments not always can be called for; if any part-of-speech can function as a predicate; if observation of language tells us that subjects are frequently non-existent in clauses, so that one cannot define a predicate as an attribute to an entity represented within the limits of the clause, or, more generally, as depending on a subject — then how do we know what a predicate might be and, consequently, how can we define a clause?

Before getting into the analysis of unipartite clauses, a few words on the interface between syntax, discourse, information structure and prosody are in order.

4 Syntax, discourse, information structure and prosody

In addition to the general premises set above (§2), I build on more specific premises as regards syntax, discourse, information structure and prosody:

- The syntactic approach adopted here is functional, communicational, discursive and information oriented. As such, syntactic components take their conceptual status from a complex analysis of which the primary originating force is contextual.
- Syntax, information structure and prosody integrate in spoken language structure, forming a coherent unity.
- Prosody is a formal feature of spoken language no less than segmental features.
- Prosody is the main tool we use for spoken language segmentation.
- For the recipient, prosody is the lead to reach a correct interpretation of the segmental structure and consequently a sound interpretation of the information conveyed.

From the recipient's perspective, prosody is a *sine qua non* when trying to delimit units of spoken language (METTOUCHI *et al.*, 2007; IZRE'EL; SILBER-VAROD, 2009). Prosodic units encapsulate corresponding segmental units, which — together with their suprasegmental features — constitute information units. Information units in themselves can either overlap or interface with syntactic units. As our concern here is with basic clause structure, it will suffice to define two units in the prosodic hierarchy: *prosodic module* and *prosodic set*.

Prosodic module (henceforth: PM; aka “intonation unit”, “tone group”, “prosodic group”, or the like), which has been determined as having a coherent intonation contour (CHAFE, 1994, p. 57-60), encapsulates a segmental unit of language to be termed *segmental module*, forming together an *information module* (IM) (cf. TAO, 1996, §§9.1-2 for what he terms *speech units*). The boundaries of IMs are therefore defined by prosody. There are two main classes of boundaries: major (which indicates terminality) or minor (which indicates continuity). Both are indicated by their respective boundary tones. A major boundary is also the boundary of a *prosodic set*.

Prosodic set is defined as a stretch of speech ending – as its default manifestation – in a major boundary. A *prosodic set* can consist of one or more PMs of which the last one ends in a major boundary, whereas any (optional) previous PM ends in a minor boundary.⁶

Whereas a PM encapsulates a *segmental unit*, forming together with it an *information module* (IM), a *prosodic set* encapsulates an *information set* or an *utterance* (cf. MONEGLIA, 2005, §1.2). I take the *utterance* to be the basic discourse unit of spontaneous spoken language (IZRE'EL, forthcoming).

As regards syntax, it is suggested that the utterance is the default domain of the clause, whether it consists of a single IM or more. The utterance is the biggest information unit that can contain a clause. A clause cannot spread beyond the boundaries of a single utterance. In other words, a major prosodic boundary indicates the terminal boundary of a clause. When an utterance consists of more than a single clause, a clause can be encapsulated by a PM. An IM can consist of either a phrase, being a component of a clause, or of a complete clause. An utterance can include additional elements to a clause or consist of a *clause set*, or, rather, a *spoken sentence*; i.e., two or more clauses joined together, thus conveying a single, integrated message. An utterance can therefore be regarded as the domain of a clause set (consisting of a single clause or more) or a spoken sentence. Thus, a sentence — like a clause — is delineated by an utterance. The interface between prosodic and segmental units can be outlined as follows:

Prosodic units	Discourse units	Syntactic Units
Prosodic Module (PM) (one of two or more in a Prosodic Set)	Information Module (IM) (one of two or more in a an utterance)	Phrase / Clause (/ Spoken sentence)
Prosodic Set	Utterance	Clause / Spoken sentence

For further details see IZRE'EL, forthcoming.

⁶ For some exclusions and a more comprehensive study of these units, see IZRE'EL, forthcoming. See also below, note 12.

5 What is a clause? What is a predicate?

Like many recent approaches to clause structure, I take the predicate to be its core component. As mentioned, I do not regard arguments as necessary components within the syntactic structure. Therefore, the predicate is the only necessary component — and a sufficient one — to constitute a clause. In other words, a clause is defined as a syntactic unit consisting minimally of a predicate. A predicate can be either nuclear or extended; in other words, it can consist of either a single element (phrase, word or part of a word) or be seen as a domain. Since any part of speech can function as predicate; since the predicate cannot be defined as an attribute to an entity represented within the limits of the clause, or, more generally, as depending upon a subject; and since it need not be related to any arguments — a new perspective of what consists of a predicate is in order. As mentioned, a discourse-related approach is taken.

The predicate (or the predicate domain) is viewed as the element carrying the informational load of the clause within the discourse context, which by default will include a newly introduced element (cf. CHAFE, 1994, p. 108). By default, the focus of the clause will be found within the predicate domain. Essentially, the predicate carries the modality of the clause.

As taken here, modality is the means by which a proposition can be actualized. This view of modality as an inherent, indispensable characteristic of the clause, basically follows the path of francophone linguistic schools (BALLY, 1965, §§28, p. 46-49, 51-54; LE GOFFIC, 1993, ch. 4; LEFEUVRE, 1999, ch. 1; GOSSELIN, 2010; convenient surveys can be found in VION, 2001; JOHANSSON; SUOMELA-SAHNI, 2011).

As nicely put by Bally,

modality is the soul of the sentence; just as thought, modality is mainly realized through the action of the speaking subject. Therefore one cannot attribute the value of a sentence to an utterance unless one has discovered the expression of modality of the utterance. (BALLY, 1965, §28; translation by JOHANSSON; SUOMELA-SAHNI, 2011, p. 95)

Arrivé, Gadet & Galimiche suggest the following guidelines for the concept of modality:

1. On a strictly logical level (modal logic), *modality* is symbolized by a system comprising two values: *possibility* and *necessity*. ...
It is convenient ... to make a distinction between *epistemic* modalities and *deontic* modalities. ...
2. *Modality* defines the status of the sentence, taking account of the attitude of the speaking subject with regard to his utterance and the addressee. Generally, distinction is made between modalities of assertion (which in itself divides into affirmation and negation), interrogation, exclamation and command. Modalities can combine: a sentence can be both interrogative and negative (...), imperative and exclamative. But not all combinations are possible: there is necessarily a contradiction between affirmation and negation. (ARRIVÉ; GADET; GALIMCHE, 1986, p. 390; my translation).

As noted by Nuyts (2005a), the notion of modality is best viewed as a supercategory, since “the domain is usually characterized by referring to a set of more specific notions, each of which is defined separately, and which may be taken to share certain features motivating their grouping together under the label *modality*, but which differ in many other respects” (NUYTS, 2005b, p. 1). With this in mind, one will recall the use by some authors of the plural *modalities* (French: *modalités*), or *modality variants* (e.g., GOSSELIN, 2010; MARTIN, 2015, 68ff.). As noted by Kiefer,

[t]hree major approaches [to modality] can be distinguished.
(i) Modality is related to necessity and possibility, it is used to relativize the validity of propositions to a set of possible worlds. On this view, modality is not necessarily propositional, it may also include nonpropositional aspects of the sentence. (ii) Any modification of a proposition comes under the heading of modality. According to this view, volitional, emotive, evaluative modifications, too, belong to modality, in spite of the fact that these modifications are not related to necessity and possibility. (iii) Modality is what the speaker is doing with a proposition. This notion of modality includes (i) and (ii): in addition, it also covers illocution, in particular, the speech acts of imposing obligation and granting permission. (KIEFER, 2009, p. 179)

The approach taken here is indeed rather comprehensive and closer to Kiefer’s option iii. Modality has thus a much broader scope than

it is usually conceived by other schools, notably Anglo-Saxon linguistic schools (e.g., PALMER, 2001; BUTLER, 2003, ch. 9), and includes not only the commonly known, consensual types of epistemic, evidential, deontic, dynamic and their like, but also assertion (pace NARROG, 2005, §2.3.1; HACQUARD, 2011, p. 1484, among many others), polarity (cf. HALLIDAY, 2014, §4.5; BUTLER, 2003, ch. 9), and beyond. It further includes sentence (or clause) modalities as used in francophone linguistic schools, specified above. A wider perception of modality has been suggested also in non-francophone linguistics schools, including Anglo-Saxon ones. For Fillmore, modality is the non-propositional component of a clause, thus including tense, aspect, mood and negation (FILLMORE, 1968, p. 23-24). Using a more restricted view of modality, Frajzyngier (1985, 1987; FRAJZYNGIER; SHAY, 2016) defines *clause* as

the smallest formal unit that has a modal value, such as ‘assertion’, ‘negation’, ‘question’, ‘hypothetical’, etc., depending on what kinds of modalities are encoded in a given language. The expression ‘having modal value’ does not mean that the unit itself codes modality. In many languages there is an unmarked modality, which is usually the assertive or affirmative modality (FRAJZYNGIER; SHAY, 2016, p. 179).

For spoken language, prosody will be regarded as basic for modality *signata*. Already Bally claimed that prosody (for him: *intonation*) is primary among non-articulatory elements that can enable the production of a sentence. For Bally, “every sentence is pronounced with an autonomous intonation that corresponds to the nature of thought” (BALLY, 1965, §50; my translation). Of course, prosody is not the only means by which modality is being represented, although it seems to be a basic one (for French see LE GOFFIC, 1993, §§51-59; MARTIN, 2009, p. 86-92; 2015, p. 68-75). According to Martin,

[t]he prosodic structure being assumed (...) independent from the sentence text modality (i.e. the one possibly indicated in the text itself) is correlated with a modality without direct relation with other modality (syntactic, morphologic) markers eventually present in the sentence. (MARTIN, 2015, p. 68)

While there is a lot more to say about modality and its forms, for our needs here suffice is to say that prosody and modality are linked


so as to enable us to see the inherent bond between clause structure and prosody.

By default, the predicate carries with it assertive (or declarative) modality. The traditional notion of assertion has always been central to the definition of predication (GOLDENBERG, 1998, p. 156-157). The thesis advanced here is that a unipartite clause does not lean on a subject. Therefore, the load of assertion (at least in unipartite clauses) is carried exclusively by the predicate domain. The same can be said of other types of modality as it is conceived here, and indeed of modality in its entire gamut.

It will be noted at this juncture that one must distinguish between semantic or pragmatic levels and the syntactic level, which is the formal means language uses to represent meaning. As we have seen, every part-of-speech can become a predicate, so that a formal definition according to segmental features seems irrelevant, especially in unipartite clauses. The main formal features used for detecting a predicate (or a predicate domain) are therefore suprasegmental, notably segmentation, final prosodic contour, and accents. For example, utterances consisting of only a single word can be defined as predicates, and hence complete clauses, using prosodic criteria (see examples in §6 below), although informational features (message, new information) will be present as well. Other elementary examples are: basic declarative modality will by default be indicated by a final fall (MARTIN, 2015, p. 72; IZRE'EL, forthcoming); focus will be marked by prosodic accent, although segmental means can also mark focus. In any case, the terminology used here, viz., *predicate* and *subject*, are essentially syntactic, albeit their interrelationship with semantic and pragmatic notions.

6 Unipartite clauses

As mentioned, a unipartite clause is a clause that consists of only a predicate domain. Ex. 9 exhibits some typical unipartite clauses:

- (9) [1] sp2: 'moruʃ ||
Morush
'Morush,'⁷ 
- [2] sp1: ma 'motek ||
what sweetie
'What, sweetie?'
- [3] sp2: arba'a ja'mim |
four days
'(For) four days - '
- [4] 'ʃva = meot 'ʃekel le = 'zug ||
seven=hundreds shekel to=couple
'(the cost is) seven hundred shekels for a couple.'
- [5] sp1: bli 'kesef ||
without money
'(This is) very cheap.'
- [6] sp2: na'χon /
right
'Isn't that so?'
- [7] sp1: 'ejfo /
where
'Where?'
- [8] sp2: be = 'holidej in ha = χa'daʃ ||
in=Holiday Inn DEF=new
'At the new Holiday Inn.'
- [9] sp1: daʃ ||
enough
'Wow!'
(OCD_2_sp2_059-063; sp1_027-030)

⁷ In *CoSIH*, personal names (in this case, a nickname) have been changed in transcription and eliminated in sound for privacy. In the sound files, names have been replaced by the actual pitch contour, produced by Praat <<http://www.fon.hum.uva.nl/praat/>>.

In this exchange, none of the utterances conforms to the traditional view of *clause* as a unit consisting of both subject and predicate and therefore capable of being analyzed in terms of what is usually regarded as a canonical clause. However, each of the utterances in lines [1], [2], [5], [6], [7], [8], [9] (which in this case each consists of a single IM) meets the requirements of the definition of a predicate as suggested above (§5) and thus constitutes a (unipartite) clause, conveying new information and carrying modality: vocative (IM [1]),⁸ interrogative (IMs [2],⁹ [6], [7]), assertive (IMs [5], [8]), or exclamative (IM [9]). Also, all units have a focus indicated by prosodic features. IMs [3]-[4] make an interesting case. IM [3] recalls a short exchange regarding a weekend at a hotel which took place almost two minutes before returning to this issue here. At this point in the conversation, it is invoked not by repeating the exact phrase used before ('weekend') but by indicating the time span of the hotel stay, viz., 'four days'. Therefore, this IM seems to introduce a new piece of information into the discourse. The modality carried by this phrase is somewhat obscured by the minor boundary tone. Had it been a major

⁸ Vocatives pose difficulties for syntactic analysis (SONNENHAUSER; AZIZ HANNA, 2013). At times, they are being referred to as "extragrammatical" (e.g., DANIEL; SPENCER, 2009; for English vocatives see BIBER *et al.*, 1999, §14.4.1; HALLIDAY, 2014, §4.3.4, who describes vocatives as outside the scope of the Mood system; CARTER; MCCARTHY, 2006, §§116-118). That an address or calling attention like 'Jack!' or 'Sir!' should be regarded as modal will be clear if we realize that it is in fact a *request* or an *order* to pay attention. If an address like these ones forms an entire utterance or comprises in itself an IM, it would carry its own independent intonation contour, forming an independent PM. In such cases, the intonation contour will be observed as indicating the modality of the IM. Of course, such an IM carries informational load with it; if it forms a separate PM it will usually be focused; and in some cases it will manifest "newness" of the address form in terms of the discourse flow (cf. CHAFE, 1994, ch. 9). Chafe has observed that "a substantive intonation unit usually (*but not always*) conveys some new information" (p. 108; my emphasis; for substantive and regulatory units see n. 2 above). While Chafe has limited this observation to substantive units, the general analysis suggested here will be valid for many regulatory units as well, although not to all of them. The behavior of these two different types of units should be subject to further investigation (cf. TAO, 1996, p. 59). In any case, vocatives such as the one discussed here may well be regarded as unipartite clauses.

⁹ The predicate is the interrogative pronoun *ma*. Unlike the vocative in IM [1], the additional element does not conform to the requirements of constituting a predicate and is taken to be external to the clausal structure (cf. IZRE'EL, forthcoming, §3.5.1).

boundary tone, there would be no doubt about the assertion expressed by this IM, making it a clear declarative clause, meaning something like ‘(It is) four days’ or ‘(We have) four days (at the hotel).’ Nevertheless, the prosodic contour — with an especially strong accent on *jamim* ‘days’ — may well be seen as a modality signal. The minor boundary tone, which indicates continuity, is needed for signaling the link between this IM ([3]) and the following one (IM [4]), which in itself unmistakably conforms to the criteria suggested above for a unipartite clause.

It will be recalled that each utterance, being a stretch of speech encapsulated by a prosodic set, is by definition delimited by a major prosodic boundary, which accordingly indicates its terminal point. As such, an utterance is the largest discourse unit that can contain either a single clause or (in the case of IMs [3]-[4] in Ex. 9) a clause set (=a spoken sentence; IZRE’EL, forthcoming, §3.6). Looking at it from a different angle, a major prosodic boundary always indicates the end of a clause and therefore also the beginning of a new clause in the following utterance (prosodic set). As it is exemplified in Ex. 9, each utterance includes a predicate domain which carries the informational load of the clause within the discourse context; each includes a newly introduced element; all units are focused via prosody; and each one carries the modality of the clause, again, indicated by prosody.

In Ex. 10, the speaker tells a piece of gossip about a couple who takes breaks during working hours:

(10) [1] at mari'χa et = ha = 'reaχ ʃel = ha = ʃam'po ||
 you.SGF smell.PTCP.SGF ACC=DEF=smell of=DEF=shampoo
 ‘You smell the shampoo.’




[2] mi = ʃne'hem ||
 from=both.of.them
 ‘From both of them.’
 (OCD: 41’:32.5”-41’:35.2”)

PM [1] ends in a major prosodic boundary and forms an IM that constitutes a complete clause; IM [2] includes what is usually regarded as an “afterthought”. Prima facie, the term “afterthought” implies only that a stretch of speech follows another one, and seems not to differ from “right dislocation”, which seems to imply the same. However, Ziv & Grosz (1994, §2) have suggested that an “afterthought” and “right

dislocation” differ in function and in some formal characteristics, noting that an “afterthought” comes after a prosodic boundary¹⁰ and comprises a separate utterance. Decades before, Bally (1965, §§75), relying on the prosodic structure of such sequences, suggested that the two parts are autonomous, and compares them to coordinate sentences (§§102-103), very much like inserts (§§70,86). In an older article, he insists to call such units “sentences” (“j’insiste sur le mot «phrase»”; BALLY, 1941, p. 40-41). As we see in our Ex. 10, a major prosodic boundary indeed separates between the two speech stretches, thus forming two distinct utterances. Complying with the requirements for informativeness, newness, focusing and modality (assertive or declarative in this case), the prepositional phrase *mifnehem* ‘from both of them’ in IM [2], standing as an utterance on its own, will be regarded from the syntactical point of view as a predicate constituting a unipartite clause. Looking at it from the point of view of parts-of-speech classification, the structure of the word that constitutes this clause is one that will be defined as an adverbial phrase. Taking this point of view, as well as looking at the semantic structure of the utterances in both IM [1] and IM [2], one can see that the utterance *mifnehem* ‘from both of them’ in IM [2] is structurally related to the predicate nucleus *marixa* ‘smell’ in IM [1]. Of course, a virtual syntactic link between the predicate in IM [1] and the adverbial phrase in IM [2] can also be deduced, one that can be tested had the two occurred within the boundaries of a single IM (or clause). In that case, the adverbial phrase would not be regarded as a predicate of a new clause but as an adjunct, since it would not carry its own modality. One should recall that in Hebrew, one will find in the predicate position any part of speech, including prepositional phrases (see above, §3; for adverbial clauses as independent sentences see TESNIÈRE, 1966, 2015, ch. 77). In the framework proffered here, where prosody is taken as the basis for segmentation of both discourse and syntactic units, as well as on the basis of the analysis advanced above where the adverbial phrase *mifnehem* is taken to be a predicate, the relationship between the two utterances must be seen on an inter-sentential level (cf. MITHUN, 2005).

¹⁰ Ziv & Grosz claim that an “afterthought” follows a pause. As pause is not a necessary requirement of prosodic boundary (AMIR, SILBER-VAROD; IZRE’EL, 2004), I would rather rephrase this requirement to mean a prosodic boundary, probably a major one, as is the case here.

It will be noted, that not all defining features will always be present in a clause. In Ex. 11, a team of the civil guard are about to take off from their base.

- (11)[1] sp1: tsa'riχ lik'not ʃti'ja ||
 need to.buy drink
 'We have to buy drinks.' 
- [2] sp4: tik'ne ba'dereχ ||
 2SGM.buy in.the.way
 'Buy (them) on the way.'
- [3] nu /
 come.on
 'Come on!'
- [4] sp1: tov ||
 good
 'Okay.'
 ...
- [5] ani = ro'tse lik'not ga'dol ||
 I=want to.buy big
 'I want to buy a big (bottle).'
- [6] sp2: ga'dol /
 big
 '(A) big (one)?'
- [7] sp1: ga'dol ||
 big
 '(A) big (one).'
- (P311_2_sp1_398-404; sp4_105-106; sp2_126)


Sp1 says that he wants to buy a big bottle of soft drink (IM [5]), introducing the component 'big' into the discourse. Therefore, when sp2 asks a verification question, the adjective *gadol* 'big' (IM [6]) is no longer new. What is new is the interrogative modality, indicated by prosody. When sp1 repeats it, the modality is again assertive, as it is in IM [5].

In this case, the defining feature of newness is not fulfilled. Still, all other three features for defining IM [7] as predicate and clause are there: informativeness, (assertive or declarative) modality, and focus, signaled by prosody.¹¹ It should be noted, that both IMs are delimited each by major prosodic boundaries, thus constituting each an utterance. It will be recalled (§4) that an utterance is the domain of a clause. It should be emphasized, that a major boundary does NOT define a syntactic unit but an informational one, although prosody has a role also in the definition of predicate in that it may signal modality and focus. Predicates, and by consequence also clauses, are defined independently from utterances, albeit their interface and their correlation at the utterance terminal boundary.

Every discourse takes place in a specific location, occurs at a specific time, and has its direct interlocutors, indicated in the discourse by the first and second personal pronouns. This is the point of departure for all deixis, the *origo* ('origin'), to use Karl Bühler's (1934) term (ABRAHAM, 2011, p. xviii). An intricate system of means is used to refer to elements in the conceptual world by linguistic signs, whether such elements are external to the discourse or occurring within it. Discourse structure uses a variety of deictic and anaphoric elements to refer to these items, notably when reference recurs in the discourse. Recurrent reference may be called for by reduced referential expressions (e.g., independent pronouns, pronominal clitics or affixes) or may not be explicitly made at all. As mentioned above (§1), there are many languages which systematically avoid the use of referential expressions (see further KIBRIK, 2011, ch. 3). Within the boundaries of a clause, reference can be made in either the subject position or in the predicative domain or in both. Of course, our interest here lies with clauses where no subject is present. We shall see that unipartite clauses are not dependent on referential representation at the subject position.

In Ex. 12, a military commander (sp3) notices a telephone ringing while reciting instructions during a roll-call of his soldiers:

¹¹ Sp1 utters this utterance in an unnatural sound and prosodic contour, which seem to convey some sort of ridicule.

- (12)[1] sp3: ʃel= 'mi ha= |
of=who DEF=
‘Whose is the’ 
- [2] 'pelefon /
cellphone
‘cellphone?’
- [3] spX: ets'l= i ||
at=1SG.GEN
‘(It is) with me.’
- [4] b = a = 'tik ||
in=DEF=bag
‘In my bag.’
- (P423_1_sp3_005-006; spX_001-002)

One of the soldiers responses first by saying *etsli* ‘(it is) with me’ (IM [3]), then by complementing it by specifying where exactly the cellphone is (IM [4]), probably making an excuse as to why he was not aware of its being there or its being turned on. In any case, the first clause (IM [3]) illustrates a predicative use of the complex *etsli* ‘with me’ in an utterance constituting a unipartite clause, hence a predicate domain. Obviously, the pronominal clitic is the nucleus of the predicative domain, being the core of information given. The following IM also constitutes a complete utterance, being delimited by two major boundaries. This utterance too can be defined, by its own characteristics, as a predicate, and therefore as a clause: it communicates new information, it carries declarative modality, and the focus is indicated by the prosodic accent, which in this case correlates with the only word-stress found in this utterance, constituting of a single prosodic word, yet in a higher pitch and intensity than the expected ones, very much like the preceding one-word utterance, *etsli*. The anchor for both predicates is ‘cellphone’, mentioned previously by sp3. Note, however, that neither the clause in IM [3] nor the one in IM [4] has any structural relation (i.e., on the formal level) to the referential element *pelefon* ‘cellphone’, which, in any case, will not be regarded as subject for neither clause.

Many unipartite clauses are anchored in a previous discourse, notably in adjacent utterances, like questions and answers (see, *inter alia*, CULICOVER; JACKENDOFF, 2005, ch. 7; GINZBURG, 2012, ch. 7). These are, however, only a part of the variety of occurrences of unipartite clauses. Givón (1992 [=2017, ch. 2]; cf. 2001, v. I, §9.5) has shown a significant correlation between the occurrence of clauses without representation of the referent and referential distance, i.e., the gap between the current and previous representation of the referent in the discourse. From data collected in several languages, Givón shows that the mean distribution of clauses without an explicit representation of the referent (for him: “zero anaphora”) will reach up to 100% of the occurrences when they immediately follow a referential representation in a previous clause. On the other hand, referents tend to be overtly and explicitly represented in the discourse the larger the gap from a previous occurrence of the same referent becomes (see his table in GIVÓN, 1992, p. 21 [=2017, p. 45]). For a more complex view of referential choice see KIBRIK, 2011, part IV.

I have mentioned above (§2) that corpus data reflect the perceived language rather than the produced one. An interesting case showing the gap between the respective speaker’s and hearer’s grounds for communicative exchange is the excerpt presented as Ex. 13. Sp1 tells her interlocutor, sp2, about her forthcoming trip to Thailand, resulting in this short exchange:

- (13) sp1: ‘In a short while I am in Thailand.’
 sp2: ‘You didn’t mention it. When are you leaving?’
 sp1: ‘29th of July.’

Sp2 does not continue to enquire about the trip, and she says instead:

lo na'im li miske'na ||
 NEG pleasant to.me poor.PTCP.SGF
 ‘I feel uncomfortable; poor her (la pauvre!).’
 (Y111_sp2_154)



And she continues:

‘And well ... And what will they do? And what will they do?’

Sp1 does not understand and asks:

‘What is it that you feel uncomfortable about?’

There follows a side-talk, during which sp1 goes to prepare herself some coffee, and when she returns, she asks again:

‘What is it that you feel uncomfortable about?’

Sp2 responds:

hi hal'χ-a ha'bajta ||
she went-3SGF homeward
‘She went home (i.e., got fired).’
(Y111_sp2_158)




Sp1 finally understands that her interlocutor was speaking about a colleague who had been fired from work:


‘Yes. I know. I discovered it yesterday when she said goodbye.’

This exchange shows the difference in active memory between participants in the conversation and therefore the capability of anchoring. Whereas the referent for the adjective *miskena* ‘poor.SGF’ is found in the active memory of sp2, it is inactive in the memory of sp1. Whereas for sp2 the predicate *miskena* ‘poor.SGF’ is anchored to an extra-*origo* referent, for the recipient this unipartite clause is unanchored, so that she has to ask for explanation. Interestingly, when sp2 helps her by making the reference, she does not use a full reference but a reduced one (i.e., the pronoun *hi* ‘she’) which seems enough for sp1 to indicate to sp2 that the referent has now been raised to her active memory.

In Ex. 14, the speakers are arriving in a place that they had not visited for a long time and try to locate the house. Following the request of sp1, who is the car driver, sp2 introduces a sign that will help the driver to find the place:

- (14) sp1: ‘Remind me where is the house.’
 ...
 sp2: jeʃ kazot e | kni'sa le =χana'ja ||
 EXT like.this uh entrance to=g garage 
 ‘There is a sort of entrance to a garage.’
 (Sh5: 2h:59':50"-53")

The existential particle *jeʃ* is traditionally analyzed as predicate in all contexts (GLINERT, 1989, §16.9; SCHWARZWALD, 2001, p. 96; KUZAR, 2012, §155; ZIV, 2013). However, it is rather the new referent introduced into the discourse that is to be regarded as a predicate. In this and many similar contexts, the existential particle *jeʃ* is better viewed as a presentational particle, although without stripping it of its existential meaning (cf. JESPERSEN, 1924, p. 154-6; MCNALLY, 2011, p. 1833; among many others). Compare the use of the presentation particle *hine* ‘here, now’ in Ex. 15:


- (15) 'hine setʃu'an |
 PRES Sichuan 
 ‘Here (is) Sichuan,’
 (OCh_sp1_027)

Here, the speaker looks at an atlas and finds sites he had visited while he was visiting China. In both cases, the existential particle (Ex. 14) or the presentation particle (Ex. 15) introduce new element into the discourse. In both cases, all other criteria for establishing these phrases as predicates are also present.

Indeed, there are cases where either the existential particle or the presentational one will be regarded a predicate. This will be the case where the other component in the clause, the so-called pivot, will be given. In such cases, the focus will be on the respective particle rather than on the pivot. With the analysis given here, the uses and functions (presentational, existential, locative, possessive, etc.) of the particle *jeʃ* and related forms (notably its negative counterpart *ejn*) should be subject for further research (IZRE'EL, in preparation). In any case, the type of presentational-existential clause represented in IM [2] of Ex. 14 should be regarded as a unipartite clause.

There are cases where the predicate cannot be shown to have an anchor in elements that have explicit linguistic expression in the discourse; rather they are anchored in elements that are external to the discourse, either within the *origo* of this specific discourse or external to it (cf. GIVÓN, 1992, §6 [=2017, ch. 2, §4]).

In Ex. 16, the speaker interrupts the flow of the conversation, feeling that something went wrong with his recording mission. He utters:

- (16) ha = hakla'tot = jeli ||
 DEF=recordings=my 
 'My recordings.'
 (P423_2_sp1_433)

In this example, the predicate 'my recordings' has no previous or any other reference in the discourse. Rather, it refers to a situation in the physical world, in this case within the *origo*, where even the situation as felt by the speaker remains unmentioned.

Finally, there are predicates that are neither anchored in the discourse at all nor do they have any obvious, direct anchors – either internal or external. The most conspicuous case of unanchored clauses are those introducing a brand new topic – or referent – into the discourse via a presentational construction (cf., *inter alia*, LAMBRECHT, 1994, §4.4). One way of introduction brand-new referents into the discourse in Hebrew is by using the existential particle *jef* (cf. Ex. 14), as in Ex. 17:

- (17)[1] tiʃme'u da'var ||
 hear.PL thing 
 'Listen to this:'
- [2] jef ma'kom |
 EXT place
 'There is a (certain) place'
- [3] ber'χov |
 in.street
- [4] le'vinski |
 Levinsky
 'in Levinsky Street'

- [5] be |
in
- [6] tela'viv |
Tel.Aviv
'in Tel-Aviv;'
- [7] 'mi'fei |
someone.SGF
- [8] fe |
that
- [9] o'sa |
make.SGF
- [10] tavli'nim |
spices
'(There is) someone (there) who makes spices,'
- [11] fe |
that
'who'
- [12] ro'kaχat |
concoct.SGF
'concocts ...'
- [13] lo ro'kaχat ||
NEG concoct.SGF
'not concocts,'
- [14] be'etsem marki'va ||
in fact put.together.SGF
'in fact, combines.'
- [15] kol = mi'nej |
all=sorts.of
'all kinds of'
- [16] e |
uh

- [17] tamhi'lim |
blends
'blends'
- [18] šel = kol = mi'nej tavli'nim be'jaħad ||
of=all=sorts.of spices together
'sorts of uh combinations of various kinds of spices together.'
(Sh2c: 38':35.3"-38':49.6")

Following a discourse-regulative comment (SG [1]), the speaker introduces her new topic by an existential clause (IMs [2]-[6]). As mentioned above for Ex. 14, the existential particle *ješ* should not be regarded in such contexts as predicate but rather as sort of a presentational particle. Thus, the existential clause in IMs [2]-[6] will be regarded as a unipartite clause, and since it introduces a brand new topic into the discourse, it will be classified as unanchored.

In this excerpt, after the initial reference to 'a place in Levinsky street in Tel-Aviv' is made, the speaker introduces another referent, this time not making use of the existential particle, perhaps because now the newly referential expression is anchored in the already presented location (IMs [7]-[18]). The utterance in IMs [7]-[18] is an expanded unipartite clause, which includes two subordinate clauses which are unipartite clauses all the same (IMs [9]-[10]; [12]-[18]), each embedded by the element *še* 'that' (IM [8], IM [11]) with an inserted parenthesis (IMs [12]-[14]).¹²

A preliminary, illustrative classification of predicates in unipartite clause, aiming at establishing their relational position in a linguistic or extra-linguistic context, has been offered in Izre'el (2018, §4).

7 On wheelless automobiles and one-room houses

We have started our endeavor to find out a different approach to clause structure because the gap between grammatical tradition and authentic linguistic data was too large to embrace (§1). Some discomfort from the allegedly safe, paved path of tradition has sometimes been

¹² Parenthetical utterances may interfere the sequence of a running utterance (IZRE'EL; METTOUCHI, 2015, §3.3). They can end in a major boundary, which, in such cases, does not mark the end of the matrix utterance (IZRE'EL, forthcoming, §3.7.2.1).

expressed. I have already cited (§1) Iwasaki's reservations regarding the use of the terms "ellipsis" and "zero anaphora" in the context of Japanese linguistics. For Kibrik (2011, p. 44), "zeroes are not a theoretical construct but rather a convention of representation." Nariyama (2007), trying "to bring more viable treatment of ellipsis particularly for NLP applications", suggests an opposite way to look at "ellipsis":

[E]llipsis can be viewed as any unexpressed information that can be drawn from context, This from the perspective of production means that any information that is inferable is made into ellipsis. (§3)

[I]t is not that sentences are produced with ellipsis, but rather those words/information that are not retrievable from contexts are being verbalized. (§3.2)

'Zero' form can mean one of two implications; 1) when something is *Understandable without saying* it is because it is anaphoric, inferable, default, or the identity is known from verbal semantics, context, situational/mutual/world knowledge, non-existent or uncertain of the existence, or 2) *no such slot exists*. (§4.1)

In generation, what should be made overt are those that are required by the syntax of a language, and are not understandable without, or for a special effect, so that known information is made overt generally when there is focus/emphasis, competing information in the context, signifying paragraph/story boundary, or treated as new information. (§4.4) (NARIYAMA, 2007; emphases in the original)

Similar or related views have been expressed time and again within linguistics, e.g., the already cited claim (§1) by Lee *et al.* (2009, p. 106), that "[i]f anything, overt subjects are 'additions' to English grammar." Lee *et al.* remind us of Ong's discussion of oral literature, where he compares the analysis of oral performance, genres and styles as "literature" to a description of horses as wheelless automobiles:

Imagine writing a treatise on horses (for people who have never seen a horse) which starts with the concept not of horse but of 'automobile', built on the readers' direct experience of automobiles. It proceeds to discourse on horses by always referring to them as 'wheelless automobiles', explaining to highly automobilized readers who have never seen a horse all the points of difference in an effort to excise all idea of 'automobile' out

of the concept ‘wheelless automobile’ so as to invest the term with a purely equine meaning. Instead of wheels, the wheelless automobiles have enlarged toenails called hooves; instead of headlights or perhaps rear-vision mirrors, eyes; instead of a coat of lacquer, something called hair; instead of gasoline for fuel, hay, and so on. In the end, horses are only what they are not. No matter how accurate and thorough such apophatic description, automobile-driving readers who have never seen a horse and who hear only of ‘wheelless automobiles’ would be sure to come away with a strange concept of a horse. The same is true of those who deal in terms of ‘oral literature’, that is, ‘oral writing’. You cannot without serious and disabling distortion describe a primary phenomenon by starting with a subsequent secondary phenomenon and paring away the differences. Indeed, starting backwards in this way — putting the car before the horse — you can never become aware of the real differences at all. (ONG, 1982, p. 12-13)

The idea that tradition can be a burden for linguists is not new and may find its first expressions already in the early history of linguistic observations. Back in the 2nd century CE, Sextus Empiricus expressed the following claim in his work *Against the Grammarians*:

In familiar intercourse, ordinary people will either oppose us about certain phrases or will not oppose us. And if they oppose us, they will at once correct us, so that we have good Greek from those who live ordinary lives and not from the Grammarians. And if they are not vexed but concur in the phrases we use as being clear and correct, we too shall abide by them. (SEXTUS EMPIRICUS, 1949, p. 113 *apud* WEILER, 1970, p. 143).

Interestingly, a vigorous call challenging the linguistic tradition comes from philosophy in recent times. In Wittgenstein’s *Philosophical Investigations* one reads the following scenario:

A is building with building stones: there are blocks, pillars, slabs and beams. B has to pass him the stones and to do so in the order in which A needs them. For this purpose they make use of a language consisting of the words “block”, “pillar”, “slab”, “beam”. A calls them out; B brings the stone which he has learnt to bring at such-and-such a call. — Conceive of this as a complete primitive language. (WITTGENSTEIN, [1953], 2009, 6^e, §2)

Later on, Wittgenstein discusses these forms of language:

‘... you can call “Slab!” a word and also a sentence; perhaps it could aptly be called a ‘degenerate sentence’ (...); in fact it is our ‘elliptical’ sentence. But that is surely only a shortened form of the sentence “Bring me a slab”, and there is no such sentence in example (2). — But why shouldn’t I conversely have called the sentence “Bring me a slab” a *lengthening* of the sentence “Slab!”?... Do you say the unshortened sentence to yourself? ... does ‘wanting this’ consist in thinking in some form or other a different sentence from the one you utter?’ (WITTGENSTEIN, [1953], 2009, 12^e, in §19; emphasis in the original)

At this juncture, Jespersen’s metaphor of a one-room house is worthy of mentioning:

It is, however, being more and more recognized by linguists that besides such two-member sentences as just mentioned we have one-member sentences. These may consist of one single word, e.g. “Come !” or “Splendid !” or “What ?”— or of two words, or more than two words, which then must not stand to one another in the relation of subject and predicate, e.g. “Come along ! | “A capital idea !” | “Poor little Ann !” | “What fun !” Here we must first guard against a misconception found in no less a grammarian than Sweet, who says (NEG §452) that “from a grammatical point of view these condensed sentences are hardly sentences at all, but rather something intermediate between word and sentence.” This presupposes that word and sentence are steps in one ascending hierarchy instead of belonging to two different spheres; a one-word sentence is at once a word and a sentence, just as a one-room house is from one point of view a room and from another a house, but not something between the two. (JESPERSEN, 1924, p. 306)

Looking back almost a century since these words were written, one will see irony in Jespersen’s note that “[a]n old-fashioned grammarian will feel a certain repugnance to this theory of one-member sentences” (JESPERSEN, 1924, p. 306). More recently, vacillating between syntax, semantics and pragmatics, debate over the analysis of so-called “subsenteses” or “fragments”, elliptical structures and their like has been going on especially since the outburst of generative grammar, putting aside what may be regarded as pre-structuralist statements over the nature of this type of units as forms of sentences (cf., in addition to

works already cited above, the discussions by SEGEL, 2008, §§1-3; HALL, 2009; HARNISH, 2009; with references to previous works).

The most recent attempt to challenge accepted views is Givón's *The Story of Zero* (2017), who suggests that

zero anaphora, rather than being an exotic feature of 'pro-drop', 'empty node', 'non-configurational' languages, is the most natural grammatical device for coding maximal referential continuity in human language. And that its gradual replacement by clitic pronouns, which eventually become obligatory pronominal agreement, is a natural, universal diachronic process. (GIVÓN, 2017, p. 155)

Thus, in evolutionary terms, unipartite clauses are viewed as more basic than bipartite ones. This idea too is not novel. One may cite Grace Andrew de Laguna, who, following observations of child language, suggested that

[t]he supposition that language had its beginnings in words would seem at first sight to be supported by reference to the speech of the little child. ... [W]hile the articulate utterances of the little child bear a resemblance to the words of his elders ... they are not ... true words. ... As the baby uses a word, it is ... a sentence-word. What the baby does from the beginning ... is to talk in complete, if rudimentary, sentences. ... The simple sentence-word is a complete proclamation or command or question The independence of the primitive word with respect to other words is paid for by its dependence on the practical situation. (DE LAGUNA, 1927, p. 86-91)

Similarly, more recent research argues that protolanguage capacity is not lost in modern languages. Support for this claim is brought forward, looking at linguistic traits drawn from child language before the age of two years; pidgin and creole languages; some types of aphasia; children prevented from acquiring language during the critical period; ad hoc 'homesign' systems used by deaf children with their hearing parents; and from emerging sign languages such as Nicaraguan Sign Language and Al-Sayyid Bedouin Sign Language (TALLERMAN, 2014, §3.2). A notable illustrative case for unipartite sentences as the first evolutionary stage in language emergence would be a story told by the oldest signer among the Al-Sayyid Bedouin Sign Language community, characterized

largely by one-word propositions, separated by pauses (i.e., prosodic signs) (SANDLER, 2017, p. 70-74). One may further recall the very first stages in second language acquisition, suggested to be the basic variety of language (JORDENS, 1997, p. 290).

Whereas a diachronic-evolutionary view of language may well see unipartite sentences as more primitive, it still remains to be seen whether this view can hold for synchronic analyses. Obviously, not all languages show the same tendencies (SAUVAGEOT, 1971; HAGÈGE, 1978; GIVÓN, 2017). Still, it seems that the view that bipartite sentences (or clauses) are more basic than unipartite ones needs to be challenged. In any case, one must look afresh at the view that unipartite sentences/clauses are elliptical or include empty ('zero') components. A revised analysis based on novel thinking is surely in place.

8 Conclusion

Adopting a framework of an integrative approach to the structure of spoken language that includes prosody, discourse structure, information structure and syntax, has resulted in our ability to account for what has been termed here *unipartite clauses*, syntactic units consisting of only a predicate domain, i.e., a nuclear or an extended predicate. The term *predicate* has been preferred over terms from other areas of investigation (e.g., "rheme", "comment", or the like), because I wish to adhere to the domain of syntactic level of investigation. By default, the predicate (or the predicate domain) is viewed as the element carrying the informational load of the clause within the discourse context, including a newly introduced element. By default, the focus of the clause will be found within the predicate domain. Essentially, the predicate carries the clause modality.

The research for establishing the notion of *unipartite clause* in spoken Israeli Hebrew was based on a rather small collection of data, which now forms *The Corpus of Spoken Israeli Hebrew (CoSIH)*. Further research, based on this corpus and on a larger collection of texts, will surely enhance our understanding of both the nature and the functions of unipartite clauses. It is my hope that research following the lines suggested here will be applied to other languages than Hebrew. As has already been mentioned briefly above, many other languages, spoken and written alike, attest similar structures in various degrees of frequency.

Hebrew, with its nature of constituting predicates from all parts of speech rather than confine it to verbs, has been productive to illustrate a fresh look at clause structure and the nature of predicate in spoken language in particular and in language in general (for some notes on similar structures in written Israeli Hebrew see RUBINSTEIN, 1968, ch. 6; SADKA, 1991; see further BERMAN, 1980).

Acknowledgements

The ideas suggested in this paper have been developed over a long period of time, during which I presented them before a variety of audiences, in both local and international conferences, and, especially, in graduate seminars at Tel-Aviv University. I thank all those, too numerous to be mentioned by name, from whom I have learned throughout the years. Special thanks are due to the participants in the meetings of a circle on the dawn of linguistics held in my home during 2013-2014, especially to Orna Harari, Eilam Alloni, Sol Azuelos-Atias, Avi Gat-Rimon, Anna Inbar, Akkad Izre'el, Naama Lemberg, and Maayan Liebrecht. I further thank two anonymous reviewers for their important comments and the editors for inviting me to contribute a paper to this volume and for accepting my paper for publication.

References

- ABRAHAM, Werner. Traces for Bühler's Semiotic Legacy in Modern linguistics. In: BÜHLER, Karl. *Theory of Language: The Representational Function of Language*. Translated by Donald Fraser Goodwin, in collaboration with Achim Eschbach. Amsterdam: John Benjamins, 2011. p. xiii-xxvii.
- ACKRILL, J. L. *Aristotle's Categories and De Interpretatione*. Oxford: Clarendon Press, 1961. (Clarendon Aristotle Series)
- ALLAN, Keith. Aristotle's footprints in the linguist's garden. *Language Sciences*, Elsevier, v. 26, n. 4, p. 317-342, 2004. Doi: <https://doi.org/10.1016/j.langsci.2003.05.001>

AMIR, Noam; SILBER-VAROD, Vered; IZRE'EL, Shlomo. Characteristics of Intonation Unit Boundaries in Spontaneous Spoken Hebrew: Perception and Acoustic Correlates. In: BEL, Bernard; MARLIEN, Isabelle (Ed.). INTERNATIONAL CONFERENCE SPEECH PROSODY, 2004, Nara, Japan. *Proceedings...* Nara: ISCA, 2004. p. 677-680.

APOLLONIUS DYSCOLUS. *Περὶ συντάξεως*. Apud HOUSEHOLDER, Fred W. *The Syntax of Apollonius Dyscolus*. Amsterdam: Benjamins, 1981.

APOLLONIUS DYSCOLUS. *Περὶ συντάξεως*. Apud LALLOT, Jean. Apollonius Dyscole. *De la construction (Περὶ συντάξεως)*. I-II. Paris: Librairie Philosophique J. Vrin, 1997.

ARENS, Hans. *Aristotle's Theory of Language and its Tradition*. Amsterdam Studies in the Theory and History of Linguistic Science. Amsterdam: Benjamins, 1984. (Series III: Studies in the History of Linguistics, 29)

ARISTOTLE. *Περὶ Ἑρμηνείας*. Apud ACKRILL, J. L. *Aristotle's Categories and De Interpretatione*. Oxford: Clarendon Press, 1961. (Clarendon Aristotle Series)

ARISTOTLE. *Περὶ Ἑρμηνείας*. Apud ARENS, Hans. *Aristotle's Theory of Language and its Tradition*. Amsterdam Studies in the Theory and History of Linguistic Science. Amsterdam: Benjamins, 1984. (Series III: Studies in the History of Linguistics, 29)

ARRIVÉ, Michel; GADET, Françoise; GALIMCHE, Michel. *La grammaire d'aujourd'hui: guide alphabétique de linguistique française*. Paris: Flammarion, 1986.

BALLY, Charles. Intonation et syntaxe. *Cahiers Ferdinand de Saussure*, Genève, v. 1, p. 33-42, 1941.

BALLY, Charles. *Linguistique générale et linguistique française*. Quatrième édition revue et corrigée. Berne: Éditions Francke, 1965.

BANNO, Eri; OHNO, Yutaka; SAKANE, Yoko; SHINAGAWA, Chikako. *An Integrated Course in Elementary Japanese*. Tokyo: The Japan Times, 1999.

BENAYOUN, Jean-Michel. Sujet zéro, pacte référentiel et thème. In: MERLE, Jean-Marie (Ed.). *Le sujet: Actes - augmentés de quelques articles - du Colloque Le Sujet organisé à l'Université de Provence, les 27 et 28 septembre 2001, avec le concours du CELA (EA 851) de l'UFR LAG-LEA et de l'Université de Provence*. Paris: Ophrys, 2003. p. 173-182. (Bibliothèque de Faïts de Langues)

BERMAN, Ruth. The Case of an (S)VO Language: Subjectless Constructions in Modern Hebrew. *Language*, Washington, v. 56, p. 759-776, 1980.

BIBER, Douglas; JOHANSSON, Stig; LEECH, Geoffrey; CONRAD, Susan; FINNEGAN, Edward. *Longman Grammar of Spoken and Written English*. Harlow: Longman, 1999.

BLANCHE-BENVENISTE, Claire. Les énoncés sans verbe en français parlé. Writen version of a lecture given at Naples, 23/2/2006.

BLANCHE-BENVENISTE, Claire; JEANJEAN, Colette. *Le français parlé: transcription et édition*. Paris: Érudition, 1987.

BLANK, David; ATHERTON, Catherine. The Stoic Contribution to Traditional Grammar. In: INWOOD, David (Ed.). *The Cambridge Companion to The Stoics*. Cambridge: Cambridge University Press, 2009. p. 310-327.

BÜHLER, Karl. *Sprachtheorie: die Darstellungsfunktion der Sprache*. Stuttgart: Gustav Fischer [1934] 1982. (Ullstein Taschenbuch, 1159)

BUTLER, Christopher S. *Structure and Function: A Guide to Three Major Structural-Functional Theories*. Amsterdam: John Benjamins, 2003. (Studies in Language Companion Series 63-64.)

CARTER, Ronald; MCCARTHY, Michael. *Cambridge Grammar of English: A Comprehensive Guide; Spoken and Written English Grammar and Usage*. Cambridge: Cambridge University Press, 2006.

CHAFE, Wallace. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: The University of Chicago Press, 1994.

CHOMSKY, Noam. *Syntactic Structures*. The Hague: Mouton, 1957. (Janua Linguarum, Series Minor, 4)

CHOMSKY, Noam. *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: The M.I.T. Press, 1965.

CHUNG, Karen Steffen. East Asian Linguistics. In: ALLAN, Keith (Ed.). *The Oxford Handbook of the History of Linguistics*. Oxford: Oxford University Press, 2013. p. 209-226. (Oxford Handbooks in Linguistics)

CRESTI, Emanuela. Notes on Lexical Strategy, Structural Strategies and Surface Clause Indexes in the C-ORAL-ROM Spoken Corpora. In: CRESTI, Emanuela; MONEGLIA, Massimo (Ed.). *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: John Benjamins, 2005. p. 209-256, ch. 6. (Studies in Corpus Linguistics, 15)

CULICOVER, Peter W.; JACKENDOFF, Ray. *Simpler Syntax*. New York: Oxford University Press, 2005.

DANIEL, Michael; SPENCER, Andrew. The Vocative – and Outlier Case. In: MALCHUKOV, Andrej; SPENCER, Andrew (Ed.). *The Oxford Handbook of Case*. Oxford: Oxford University Press, 2009. p. 626-634.

DE LAGUNA, Grace Andrews. *Speech: Its Function and Development*. New Haven: Yale University Press, 1927.

DIOGENES LAERTIUS. Apud LONG, A. A.; SEDLEY, D. N. *The Hellenistic Philosophers*. Cambridge: Cambridge University Press, 1987. v. I: Translations of the principal sources with philosophical commentary.

DU BOIS, John W.; CUMMING, Susanna; SCHUETZE-COBURN, Stephan; PAOLINO, Danae. *Discourse Transcription*. Santa Barbara, CA: Department of Linguistics, University of California, Santa Barbara, 1992. (Santa Barbara Papers in Linguistics, 4)

EHLICH, Konrad. Phrase averbale, phrase nominale? La constellation sémitique. In: BEHR, Irmtraud; FRANÇOIS, Jacques; LACHERET-DUJOUR, Anne; LEFEUVRE, Florence (Ed.). *Syntaxe & Sémantiques*, 6: Aux marges de la prédication. Caen: CRISCO: Centre de Recherches Inter-langues sur la Signification en Contexte; Presses Universitaires de Caen, 2005. p. 103-124.

FILLMORE, Charles J. The Case for Case. In: BACH, E.; HARMS, R.T. (Ed.). *Universals in Linguistic Theory*. London: Holt, Rinehart and Winston. [Cited from: *Form and Meaning in Language; Volume I: Papers on Semantic Roles.*]. Chicago: The University of Chicago Press, 1968. p. 21-119; references: 285-294.

FOLEY, W. A. Topic Chaining Constructions. In: BROWN, Keith (Ed.). *Encyclopedia of Language and Linguistics*. 2nd ed. Oxford: Elsevier, 2006. p. 773-777. Doi: <https://doi.org/10.1016/B0-08-044854-2/02019-8>

FOWLER, Harold N. *Plato in Twelve Volumes*. Cambridge, MA: Harvard University Press; London: William Heinemann, 1921. v. 12. Available at: <<http://data.perseus.org/texts/urn:cts:greekLit:tlg0059.tlg007.perseus-eng1>>.

FRAJZYNGIER, Zygmunt. Truth and the Indicative Sentence. *Studies in Language*, John Benjamin, v. 9, n. 2, p. 243-254, 1985.

FRAJZYNGIER, Zygmunt. Truth and Compositionality Principle: A Reply to Palmer. *Studies in Language*, John Benjamin, v. 11, n. 1, p. 211-217, 1987.

FRAJZYNGIER, Zygmunt; SHAY, Erin. *The Role of Functions in Syntax: A Unified Approach to language Theory, Description, and Typology*. Amsterdam: Benjamins, 2016. (Typological Studies in Language, 111). Doi: <https://doi.org/10.1075/tsl.111>

GENETTI, Carol. Syntax: words in combination. In: GENETTI, Carol (Ed.). *How Languages Work: An Introduction to Language and Linguistics*. Cambridge: Cambridge University Press, 2014. p. 118-149.

GIL, David. Where Does predication Come from? *Canadian Journal of Linguistics / Revue Canadienne de Linguistique*, Cambridge, v. 57, p. 303-333, 2012.

GINZBURG, Jonathan. *The Interactive Stance: Meaning for Conversation*. Oxford: Oxford University Press, 2012. Doi: <https://doi.org/10.1093/acprof:oso/9780199697922.001.0001>

GIVÓN, T. (Ed.). *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam: Benjamins, 1983. (Typological Studies in Language, 3). Doi: <https://doi.org/10.1075/tsl.3>

GIVÓN, T. The Grammar of Referential Coherence as Mental processing Instructions. *Linguistics*, de Gruyter, v. 30, p. 5-55, 1992. [Revised version: GIVÓN, T. *The Story of Zero*. Amsterdam: Benjamins, 2017. ch. 2.]

GIVÓN, T. *Syntax: An Introduction*. I-II. Amsterdam: John Benjamins, 2001. Doi: <https://doi.org/10.1075/z.syn1>

GIVÓN, T. *The Story of Zero*. Amsterdam: Benjamins, 2017.

GLINERT, Louis. *The Grammar of Modern Hebrew*. Cambridge: Cambridge University Press, 1989.

GOLDENBERG, Gideon. On Verbal Structure and the Hebrew Verb. In: _____. *Studies in Semitic Linguistics: Selected Writings*. Jerusalem: Magnes, 1998. p. 138-147.

GOSELIN, Laurent. *Les modalités en français: La validation des représentations*. Amsterdam: Chronos, 2010.

HACQUARD, Valentine. Modality. In: VON HEUSINGER, Klaus; MAIENBORN, Claudia; PORTNER, Paul (Ed.). *Semantics: An International Handbook of Natural Language Meaning*. Berlin: De Gruyter Mouton, 2011. v. 2, p. 1484-1515. (Handbooks of Linguistics and Communication Science, 33.2)

HAGÈGE, Claude. Du thème en thème en passant par le sujet: Pour une théorie cyclique. *La Linguistique*, Paris, v. 14, n. 2, p. 3-28, 1978.

HALL, Alison. Subsentential Utterances, Ellipsis, and Pragmatic Enrichment. *Pragmatics & Cognition*, John Benjamin, v. 17, p. 222-250, 2009.

HALLIDAY, M. A. K. *Halliday's Introduction to Functional Grammar*. Fourth edition revised by Christian M. I. M. Matthiessen. London; New York: Routledge, 2014.

HARNISH, Robert M. The Problem of Fragments: Two Interpretative Strategies. *Pragmatics & Cognition*, John Benjamin, v. 17, p. 251-282, 2009.

HOUSEHOLDER, Fred W. *The Syntax of Apollonius Dyscolus*. Amsterdam: Benjamins, 1981.

ILDEFONSE, Frédérique. Sujet et prédicat chez Platon, Aristote et les Stoïciens. *Archives et Documents de la Société d'Histoire et d'Épistémologie des Sciences du Langage*, Persée, Seconde série, n. 10, p. 3-34, 1994.

IWASAKI, Shoichi. *Japanese*. Revised edition. Amsterdam: Benjamins, 2013. (London Oriental and African Library, 17)

IZRE'EL, Shlomo. The Corpus of Spoken Israeli Hebrew: Textual Samples. *Leshonenu*, Jérusalem, v. 64, p. 289-314, 2002. [Hebrew]. Available at: <http://www.tau.ac.il/~izreel/publications/Texts_Leshonenu2002.pdf>.

IZRE'EL, Shlomo. Intonation Units and the Structure of Spontaneous Spoken Language: A View from Hebrew. In: CYRIL, Auran *et al.* (Ed.). *Proceedings of the IDP05 International Symposium on Discourse-Prosody Interfaces*. 2005. Available at: <http://www.tau.ac.il/~izreel/publications/IntonationUnits_IDP05.pdf>.

IZRE'EL, Shlomo. Basic Sentence Structures: A View from Spoken Israeli Hebrew. In: CADDÉO, Sandrine; ROUBAUD, Marie-Noëlle; ROQUIER, Magali; SABIO, Frédéric (Ed.). *Panser les langues avec Claire Blanche-Benveniste*. Aix-en-Provence: Presses Universitaires de Provence, 2012. p. 215-227. (Langues et Langage, 20) [A corrected version can be downloaded at: <http://www.tau.ac.il/~izreel/publications/SentenceStructure_BlancheMem.pdf>].

IZRE'EL, Shlomo. Unipartite clauses: A View from Spoken Israeli Hebrew. In: TOSCO, Mauro (Ed.). *Afroasiatic: Data and Perspectives*. Amsterdam: John Benjamins, 2018. p. 235-259. (Current Issues in Linguistic Theory, 339)

IZRE'EL, Shlomo. The Basic Unit of Spoken Language and the Interface Between Prosody, Discourse and Syntax: A View from Spontaneous Spoken Hebrew. In: IZRE'EL, Shlomo; MELLO, Heliana; PANUNZI, Alessandro; RASO, Tommaso (Ed.). *In Search of a Reference Unit for Speech: A Corpus-driven Approach*. Amsterdam: John Benjamins. Forthcoming. (Studies in Corpus Linguistics) Available at: <https://www.academia.edu/35141688/The_Basic_Unit_of_Spoken_Language_and_the_Interface_Between_Prosody_Discourse_and_Syntax_A_View_from_Spontaneous_Spoken_Hebrew>.

IZRE'EL, Shlomo. *Existential Constructions in Spoken Israeli Hebrew* (temporary title). In preparation.

IZRE'EL, Shlomo; METTOUCHI, Amina. Representation of Speech in CorpAfroAs: Transcriptional Strategies and Prosodic Units. In: METTOUCHI, Amina; VANHOVE, Martine; CAUBET, Dominique (Ed.). *Corpus-based Studies of Lesser-described Languages: The CorpAfroAs corpus of spoken AfroAsiatic languages*. Amsterdam: Benjamins, 2015. p. 13-41. (Studies in Corpus Linguistics, 68). Available at: <https://www.academia.edu/516836/Representation_of_Speech_in_CorpAfroAs_Transcriptional_Strategies_and_Prosodic_Units>

IZRE'EL, Shlomo; SILBER-VAROD, Vered. "OMER LENATEAX LENATEAX": Perception of Prosodic Groups in Spoken Hebrew. *Hebrew Linguistics*, v. 63-64, p. 13-33, 2009. [Hebrew]. Available from: <<http://cosih.com/publications/omerlenateax.pdf>>.

JESPERSEN, Otto. *The philosophy of Grammar*. London: George Allen & Unwin, [1924] 1951.

JOHANSSON, Marjut; SUOMELA-SAHNI, Eija. *Énonciation: French Pragmatic Approache(s)*. In: ZIENKOWSKI, Jan; ÖSTMAN, Jan-Ola; VERSCHUEREN, Jef (Ed.). *Discursive Pragmatics*. Amsterdam: John Benjamins, 2011. p. 71-101. (Handbook of Pragmatics Highlights, 8)

JORDENS, Peter. Introducing the Basic Variety. *Second Language Research*, Sage Journals, v. 13, n. 4, p. 289-300, 1997.

KIBRIK, Andrej A. *Reference in Discourse*. Oxford: Oxford University Press, 2011. (Oxford Studies in Typology and Linguistic Theory.)

KIEFER, Ferenc. Modality. In: BRISARD, Frank; ÖSTMAN, Jan-Ola; VERSCHUEREN, Jef (Ed.). *Grammar, Meaning and Pragmatics*. Amsterdam: John Benjamins, 2009. p. 179-207. (Handbook of Pragmatics Highlights, 5)

KUZAR, Ron. 2012, *Sentence Patterns in English and Hebrew*. Amsterdam: Benjamins, 2012. (Constructional Approaches to Language, 12)

LALLOT, Jean. Apollonius Dyscole. *De la construction (Περὶ συντάξεως)*. I-II. Paris: Librairie Philosophique J. Vrin, 1997.

LAMBRECHT, Knud. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representation of Discourse Referents*. Cambridge: Cambridge University Press, 1994. (Cambridge Studies in Linguistics, 71). Doi: <https://doi.org/10.1017/CBO9780511620607>

LEE, Namhee; MIKESELL, Joaquin, LISA, Anna Dina L.; MATES, Andrea W.; SCHUMANN, John H. *The Interactional Instinct: The Evolution and Acquisition of Language*. Oxford: Oxford University Press, 2009.

LEFEUVRE, Florence. *La phrase averbale en français*. Paris: L'Harmattan, 1999. (Collection Langue & Parole, Recherches en Sciences du Langage)

LE GOFFIC, Pierre. *Grammaire de la Phrase Française*. Paris: Hachette, 1993. (Collection Hachette Université: Langue française)

LENK, Hans. Introduction: Introduction: If Aristotle Had Spoken and Wittgenstein Known Chinese... Remarks Regarding Logic and Epistemology: a Comparison Between Classical Chinese and Some Western Approaches. In: LENK, Hans; PAUL, Gregor (Ed.). *Epistemological Issues in Classical Chinese Philosophy*. Albany: State University of New York Press, 1993. p. 1-10. (SUNY Series in Chinese Philosophy and Culture)

LINELL, P. *The Written Language Bias in Linguistics: Its Nature, Origins and Transformations*. Oxford: Routledge, 2005. Doi: <https://doi.org/10.4324/9780203342763>

LONG, A. A.; SEDLEY, D. N. *The Hellenistic Philosophers*. Cambridge: Cambridge University Press, 1987. v. I: Translations of the principal sources with philosophical commentary.

LSJ: LIDDELL, Henry George; SCOTT, Robert. *A Greek-English Lexicon*. Revised and augmented throughout by Sir Henry Stuart Jones with the assistance of Roderick McKenzie and with the cooperation of many scholars. Ninth edition, 1940; with a revised supplementary. Oxford: Clarendon Press, 1996.

MARTIN, Philippe. *Intonation du français*. Paris: Armand Colin, 2009. (Collection U: Linguistique.)

MARTIN, J.-Philippe. *The Structure of Spoken Language: Intonation in Romance*. Cambridge: Cambridge University Press, 2015. Doi: <https://doi.org/10.1017/CBO9781139566391>

MAUTHNER, Fritz. *Beiträge zu einer Kritik der Sprache*. I: Sprache und Philologie (1901); II: Zur Sprachwissenschaft (1901); III: Zur Grammatik und Logik (1902).

MAUTHNER, Fritz. *Aristotle*. Translated by Charles D. Gordon. New York: McClure, Phillips & Co., 1907. (Illustrated Cameos of Literature)

MCNALLY, Louise. Existential Sentences in English. In: MAIENBORN, C.; VON HEUSINGER, K.; PORTNER, P. (Ed.). *Semantics: An International Handbook of Natural Language Meaning*. Berlin: de Gruyter, 2011. p. 1829-1848. (Handbooks of Linguistics and Communication Science)

MERCHANT, Jason. Ellipsis: A survey of analytical approaches. In: VAN CRAENENBROECK, Jeroen; TEMMERMAN, Tanja (Ed.). *Handbook of ellipsis*. Oxford: Oxford University Press, 2015. Forthcoming in 2018. Available at: <<http://home.uchicago.edu/merchant/pubs/ellipsis.revised.pdf>>.

METTOUCHI, Amina; LACHERET-DUJOUR, Anne; SILBERVAROD, Vered; IZRE'EL, Shlomo. Only Prosody? Perception of speech segmentation. In: *Nouveaux Cahiers de Linguistique Française*, v. 28: Interfaces discours – prosodie: Actes du 2ème Symposium International and Colloque Charles Bally, 2007. p. 207-218.

MITHUN, Marianne. On the assumption of the sentence as the basic unit of syntactic structure. In: FRAJZYNGIER, Zygmunt; HODGES, Adam; ROOD, David S. (Ed.). *Linguistic Diversity and Language Theories*. Amsterdam: Benjamins, 2005. p. 169-183. (Studies in Language Companion Series, 65)

MONEGLIA, Massimo. The C-ORAL-ROM Resource. In: CRESTI, Emanuela; MONEGLIA, Massimo (Ed.). *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: John Benjamins, 2005. ch. 1, p. 1-70. (Studies in Corpus Linguistics, 15)

NARIYAMA, Shigeiko. Ellipsis and Markedness: Examining the Meaning of Ellipsis. In: INTERNATIONAL DISCOURSE ANAPHORA AND ANAPHOR RESOLUTION COLLOQUIA, 6th., Porto, 2007. *Proceedings...* Porto: Centro de Linguística da Universidade do Porto, 2007. p. 97-102.

NARROG, Heiko. On Defining Modality Again. *Language Sciences*, Elsevier, v. 27, p. 165-192, 2005. Doi: <https://doi.org/10.1016/j.langsci.2003.11.007>

NUYTS, Jan. The modal confusion. In: KLINGE, Alex; HENRIK HØEG, Müller (Ed.). *Modality: Studies in Form and Function*. London: Equinox, 2005a. p. 5-38.

NUYTS, Jan. Modality: Overview and Linguistic Issues. In: WILLIAM, Frawley (Ed.). *The Expression of Modality*. With the assistance of Erin Eschenroeder, Satah Mills and Thao Nguyen. Berlin: Mouton de Gruyter, 2005b. p. 1-26. (The Expression of Cognitive Categories [ECC], 1)

NUYTS, Jan. Surveying modality and mood: An introduction. In: NUYTS, Jan; Auwera, Johan Van Der (Ed.). *The Oxford Handbook of Modality and Mood*. Oxford: Oxford University Press, 2016. p. 1-8.

OKAMOTO, Shigeiko. *Ellipsis in Japanese Discourse*. 1985. Dissertation (Doctoral) - University of California, Berkeley, 1985. Available at: <<http://escholarship.org/uc/item/4zx1c0rg#page-1>>.

ONG, Walter J. *Orality and Literacy*. The Technologizing of the Word. (New Accents.) London: Methuen, 1982.

PALMER, F. R. *Mood and Modality*. Second edition. Cambridge: Cambridge University Press, 2001. (Cambridge Textbooks in Linguistics). Doi: <https://doi.org/10.1017/CBO9781139167178>

PLATO, *Sophist*. = PLATO. *Platonis Opera*. Ed. John Burnet. Oxford: Oxford University Press, 1903. Available at: <<http://data.perseus.org/texts/urn:cts:greekLit:tlg0059.tlg007.perseus-grcl>>.

REICH, Ingo. Ellipsis. In: MAIENBORN, C.; VON HEUSINGER, K.; PORTNER, P. (Ed.). *Semantics: An International Handbook of Natural Language Meaning*. Berlin: de Gruyter, 2011. p. 1849-1874. (Handbooks of Linguistics and Communication Science)

RUBINSTEIN, Eliezer. *The Nominal Sentence: A Study in the Syntax of Contemporary Hebrew*. Tel-Aviv: Hakibbutz Hameuchad, 1968. [Hebrew]

SADKA, Isaac. The Unipartite Utterance. In: GOSHEN-GOTTSTEIN, Moshe; MORAG, Shlomo; KOGUT, Simha (Ed.). *Shay le-Chaim Rabin: Studies on Hebrew and Other Semitic Languages Presented to Professor Chaim Rabin on the Occasion of His 75th Birthday*. Jerusalem: Akademon, 1991. p. 295-310. [= Isaac Sadka. 1997. *Studies in Hebrew Syntax and Semantics*. Beer-Sheva: Ben-Gurion University Press. 102-115] [Hebrew]

SANDLER, Wendy. What Comes First in Language Emergence? In: ENFIELD, N. J. (Ed.). *Dependencies in Language: On the Causal Ontology of Linguistic Systems*. Berlin: Language Science Press, 2017. p. 63-83. (Studies in Diversity Linguistics, 14)

SANDMANN, Manfred. *Subject and Predicate: A Contribution to the Theory of Syntax*. Second revised and enlarged edition. Heidelberg: Carl Winter Universitätverlag, 1979.

SAUVAGEOT, Aurélien. La servitude subjectale dans les langues ouraliennes. *Études Finno-Ougriennes*, Paris, v. 9, p. 15-31, 1971.

SCHWARZWALD, Ora R. *Modern Hebrew*. München: LINCOM Europa, 2001. (Languages of the World/Materials, 127)

SEGEL, Esben. Re-evaluating Zero: When Nothing Makes Sense. *SKASE Journal of Theoretical Linguistics*, Slovakia, v. 5, n. 2, p. 1-20, 2008.

SEUREN, Pieter A. M. *Western Linguistics: An Historical Introduction*. Malden, MA: Blackwell, 1998. Doi: <https://doi.org/10.1002/9781444307467>

SEXTUS EMPIRICUS. Against the Professors. In: *The Works of Sextus Empiricus*. London, 1949. (Loeb Classical Library)

SINCLAIR, John. Review of BIBER *et al.* 1999. *International Journal of Corpus Linguistics*, John Benjamins, v. 6, p. 339-359, 2001.

SONNENHAUSER, Barbara; AZIZ HANNA, Patricia Noel. Introduction: Vocative! In: _____ (Ed.). *Vocative! Addressing Between System and Performance*. Berlin: De Gruyter Mouton, 2013. p. 1-23. (Trends in Linguistics: Studies and Monographs, 261)

- STANTON, Robert J. The Pragmatics of Non-Sentences. In: HORN, Lawrence R.; WARD, Gregory (Ed.). *The Handbook of Pragmatics*. Malden, MA: Blackwell, 2004. p. 266-287. (Blackwell Handbooks in Linguistics)
- SWEET, NEG; SWEET, Henry. *A New English Grammar: Logical and Historical*. Oxford: Clarendon Press, [1892] 1898.
- TALLERMAN, Maggie. The Evolutionary Origins of Syntax. In: CARNIE, Andrew; SATO, Yosuke; SIDDIQI, Daniel (Ed.). *The Routledge Handbook of Syntax*. London: Routledge, 2014. p. 446-462. (Routledge Handbooks in Linguistics)
- TAO, Hongyin. *Units in Mandarin Conversation: Prosody, Discourse, and Grammar*. Amsterdam: John Benjamins, 1996. (Studies in Discourse and Grammar, 5). Doi: <https://doi.org/10.1075/sidag.5>
- TESNIÈRE, Lucien. *Éléments de syntaxe structurale*. Deuxième édition revue et corrigée. Paris: Klincksiek, 1966.
- TESNIÈRE, Lucien. *Elements of Structural Syntax*. Translated by Timothy Osborne and Sylvain Kahane. Amsterdam: Benjamins, 2015.
- TOGNINI-BONELLI, Elena. *Corpus Linguistics at Work*. Amsterdam: John Benjamins, 2001. (Studies in Corpus Linguistics, 6)
- TSUJIMURA, Natsuko. *An Introduction to Japanese Linguistics*. Second edition. Malden, MA: Blackwell, 2007. (Blackwell Textbooks in Linguistics, 10)
- VAN VALIN, Robert D., JR.; LAPOLLA, Randy J. *Syntax: Structure, Meaning and Function*. Cambridge: Cambridge University Press, 1997. (Cambridge Textbooks in Linguistics)
- VION, Robert. Modalités, modalisations et activités langagières. *Marges Linguistiques*, France, v. 2, p. 209-231, 2001.
- WEILER, Gershon. *Mauthner's Critique of Language*. Cambridge: University Press, 1970.
- WINKLER, S. Ellipsis. In: BROWN, Keith (Ed.). *Encyclopedia of Language and Linguistics*. 2nd Edition. Oxford: Elsevier, 2006. p. 109-113. Doi: <https://doi.org/10.1016/B0-08-044854-2/02003-4>

WITTGENSTEIN, Ludwig. *Philosophische Untersuchungen / Philosophical investigations*. Translated by G. E. M. Anscombe, P. M. S. Hacker, and Joachim Schulte. Rev. 4th ed. by P. M. S. Hacker and Joachim Schulte. Chichester: Wiley-Blackwell, [1953] 2009.

ZIEGELER, Debra. The diachrony of modality and mood. In: NUYTS, Jan; VAN DER AUWERA, Johan (Ed.). *The Oxford Handbook of Modality and Mood*. Oxford: Oxford University Press, 2006. p. 387-405.

ZIV, Yael. Existential: Modern Hebrew. In: KHAN, Geoffrey (Ed.). *Encyclopedia of Hebrew Language and Linguistics*. Brill Online, 2013. Available at: <http://referenceworks.brillonline.com/entries/encyclopedia-of-hebrew-language-and-linguistics/existential-modern-hebrew-COM_00000928>.

ZIV, Yael; GROSZ, Barbara. Right Dislocation and Attentional State. In: MITTWOCH, A.; BUCHALLA, R. (Ed.). *The Israel Association for Theoretical Linguistics: Proceedings of the 9th Annual Conference and Workshop on Discourse*. Jerusalem: Akademon, 1994. p. 184-199.