INTRODUCTION

Approaching diversity in Speech Studies
New methodologies under empirical perspectives

Giulia Bossaglia, Heliana Mello & Tommaso Raso
Universidade Federal de Minas Gerais

1. Why this special issue

At the beginning of August 2015, an important workshop organized by the LABLITA[1] lab of the University of Florence and the LEEL[2] lab of the Federal University of Minas Gerais took place at the Federal University of Minas Gerais at Belo Horizonte (Brazil).

The workshop aimed at discussing the segmentation criteria for spoken language and its unit of reference above the level of words, that is, the minimal communicative unit in speech. For almost a year, the 21 participants from the different invited teams had worked segmenting the same texts, one dialogue and one monologue extracted from the Santa Barbara Corpus of Spoken American English (Du Bois *et al.* 2000-2005), proposing their segmentation following prosodic criteria and different organizations in terms of reference units. Additionally, the participants applied their segmentation criteria to other languages as well, such as Brazilian Portuguese, French, Russian, Japanese, Hebrew, Athabaskan and Pomo languages, and other kinds of text analyses for these languages were also presented[3].

---

[1] http://lablita.dit.unifi.it/

[2] http://www.c-oral-brasil.org/

[3] For more details about the workshop and to download the presentations, see www.letras.ufmg.br/ixlablitaandivleel/

The workshop included an open call session as well, which featured different themes in the study of spoken language. This special issue of the journal publishes 17 selected papers presented to the open call for papers.

Among them, the reader will find an extensive range of themes pertaining to the study of speech through the analysis of several languages, namely Brazilian Portuguese, Italian, Spanish, French, English, Hebrew and Amerindian languages. In the following sections, we present the featured papers according to their general theme.

2.   Speech segmentation

Speech segmentation is directly approached in this volume by Mittmann & Barbosa, who responded to the call to the main theme of the workshop, i.e. the segmentation of speech in reference units. This theme is paramount for the study of speech because it is only through the proper segmentation of speech that the linguistic structure of spontaneous spoken languages can be studied. This means that without the prosodic organization of the speech flow we cannot know what the actual hierarchical organization and the real function of a given linguistic sequence are.

It is commonly recognized that the segmentation of speech cannot be made following criteria based on the structure of written language or strictly following sentential organization based on a verbal head. In speech, the linguistic "environments" in which to reconstruct linguistic relations are marked by prosody, and too often these environments do not correspond to a traditional well formed sentence. In languages such as English or the Romance languages, it is commonplace for between 25 to 40% of autonomous verbless sequences to be found (Biber *et al.* 1999; Cresti 2005; Raso & Mittmann 2012), as attested by different spoken corpora, encompassing large amounts of data compilation that follow a specific and explicit architecture and that can be exploited through computational means (McEnery & Hardie 2012). For spoken corpora we can now add a new requirement: the necessity to offer text-to-speech alignment in order to easily allow for its examination synchronously and every time it is required. This permits the necessary access to textual information and that conveyed by the acoustic signal to be paired, which is required so very frequently in the study of speech.

But when we say that prosody marks speech segmentation in units of reference or in smaller units inside a unit of reference we do not know how different prosodic devices work in order to realize this task and in which hierarchy they

operate (Raso *et al.* 2015). We can trust our perception, if that is supported by statistical validation; however, that does not eliminate disagreements, and still we must admit that we ignore or cannot be sure about which prosodic cues lead to declared perception. Mittmann & Barbosa present a very interesting and important project aimed at the study of phonetic features or configurations of features that lead to speech segmentation in autonomous or concluded units of reference and in prosodic envelopes inside the unit of reference, i.e. informative words or groups of words that, even if not autonomous and concluded, play a coherent role inside the structure of the reference unit. The results of such a project can lead to important consequences for natural language processing and to strongly supporting the implementation of resources for the study of speech. In fact, as everyone who has tried to build a resource for the study of speech knows, the compilation of spontaneous speech corpora is a tremendous task, which demands enormous effort from many people for the recording and annotation (prosodic, morphosyntactic, informational, etc.) of the collected data. The annotation is strictly dependent on the segmentation of the speech flow. Instruments that could aid on the segmentation of large amounts of spoken data would greatly improve, both qualitatively and quantitatively, the implementation of new resources.

Also focusing on speech segmentation, the paper by Lucente proposes speech segmentation criteria based on prosodic features for radio interviews, in order to pursue discourse analysis. It is especially interesting to find some overlapping on the criteria and the techniques used in both Mittmann & Barbosa's and Lucente's proposals, such as the use of V-V segmentation and SGDetector (Barbosa 2006) for the detection of stress groups. The proposal by Mittmann & Barbosa includes a wider set of parameters besides the duration one. Both the proposals compare their results with those found by perceptual manual segmentations. The previously mentioned proposals about segmentation criteria clearly show how speech studies have been strongly improved by statistical and computational methods. The competence in such methods constitutes now a necessary skill for speech scholars, as it is repeatedly demonstrated also by several other papers on different topics presented here, such as, for example, Ferrari & Teixeira, Cruz & Rilliard, Silva, and Raso & Vieira, among others.

3.   Resources for the study of spontaneous speech

The importance of spoken corpora for the study of speech is also recognized by three papers of this volume that present resources, together with several papers that base their research on existing resources.

Oliveira's paper presents the restoration and dissemination of part of a very important Brazilian corpus, the NURC Project4, historically coordinated by A. Castilho. NURC is a pioneering corpus effort developed from the end of the 1960s and still active. This corpus, besides being an important data source for the study of Brazilian Portuguese (BP), especially in a sociolinguistic dimension, constitutes today also a basis for the diachronic study of BP speech in different metropolitan areas of the country, which makes NURC one of the few resources available for this line of studies. The project coordinated by Oliveira is of extreme relevance, because it will hopefully avoid the loss of such precious recordings and will allow NURC to be used by a broad range of scholars for diverse types of studies, including segmentation and alignment that could make NURC comparable to resources produced more recently under new technologies and scientific perspectives.

The second resource presented here by Cavalcante & Ramos encompasses the annotation of 20 texts from the Santa Barbara Corpus of American English (Du Bois *et al.* 2000-2005). This minicorpus is prosodically and informationally annotated following the same criteria initially proposed by the two LABLITA-LEEL minicorpora of spoken Italian and Brazilian Portuguese (Mittmann & Raso 2011; Panunzi & Mittmann 2014), with which it is therefore comparable[5]. The Italian and the Brazilian minicorpora are extracted from the C-ORAL-ROM (Cresti & Moneglia 2005) and the C-ORAL-BRASIL (Raso & Mello 2012) corpora. These minicorpora are planned especially for the study of illocutions and information structure, and therefore their architecture is based mainly on the diaphasic variation dimension, which allows the study of a wide range of speech acts and informational structures.

The last resource presented here by Mello & Mello is a small corpus of Brazilian Portuguese as a second language spoken by Brazilian Amerindians (COLPI) from different ethnic groups and mother tongues. COLPI is the first resource of its kind to be developed in Brazil and features transcriptions and sound files which portray narratives about myths, daily life and cultural practic-

---

[4] Norma Linguística Urbana Culta (*Cultured Linguistic Urban Norm*).
[5] The Italian and the Brazilian minicorpora can be accessed at:
http://lablita.dit.unifi.it/app/dbipic/

es as reported by Native Brazilians. The documentation of BP as an Amerindian second language has both linguistic and cultural relevance and signals toward the necessity of the continuity of such efforts so that both areas can be better tended to by the academic community.

Before concluding this section, it is useful to stress the important connection between the research on speech segmentation and the compilation of corpora such as those belonging to the C-ORAL family and similar resources, for two main reasons. Firstly, because the segmentation of the speech continuum is of paramount importance for the study of spoken language at any level of linguistic analysis, as most of the papers presented in this volume clearly show. Besides, third generation spoken corpora represent the necessary condition for such research to be improved, putting the experience of corpus compilation to the service of a technologic step that can later benefit the realization of new resources. In the next two sections various papers are presented, which show on the one hand the richness and diversity of themes that adequate resources allow analyzing in spoken language, and, on the other, the emergence of promising results, which often challenge the existing literature.

4.   Corpus based studies

As previously mentioned, the remainder papers presented here are based on already existing corpora. Two papers analyze syntactic aspects of speech, mainly focusing on some aspects of subordination in spoken language.

The one by Bossaglia employs as sources the two comparable minicorpora extracted from the C-ORAL-BRASIL and the Italian C-ORAL-ROM corpora, studying the effects of rhythmic structure on the syntactic complexity of spoken Brazilian Portuguese and Italian. Rhythmic features are analyzed within the domain of a single tone/information unit, in order to better account for different behaviors of the two languages in what regards the syntax-information structure interface in the realization of complement clauses. This kind of research is possible only if the corpus is provided with prosodic information and segmentation.

The paper by Inbar studies the behavior and the possible diachronic origins of one subordinator in Hebrew, taking into account data from CoSIH (Izre'el & Rahav 2004), which is provided with a prosodic segmentation similar to that of the C-ORAL family corpora. Inbar proposes an alternative view not only of the actual uses of the subordinator in real conversations, but also of the traditional notion of subordination itself. As it is common in speech studies that are based on spoken corpora, challenging results emerge in the analysis of linguistic cate-

gories that have been defined by a long tradition of studies based only on written texts. As regards syntax, therefore, researches can greatly benefit from corpora segmented through prosodic criteria, since the reference units determined by prosody are able to show a very different behavior of spoken syntax.

Two other papers by Silva & Mello and Ferrari & Teixeira use the C-ORAL-BRASIL data to deal with two relevant aspects of Brazilian morphosyntax, respectively the motivation for the use of three different verbal negation forms and for the use of different forms of the third person subject pronoun in spoken Brazilian Portuguese. The results presented by both papers show the relevance of having available data from spontaneous speech corpora that can immensely benefit analytical efforts by offering access to the acoustic signal through the text-to-speech alignment. The understanding of the motivation for the use of different negation forms as well as pronominal forms would not be possible without the systematic access to and analysis of information conveyed by sound, especially without information conveyed by prosody. Both negation and pronominal forms in Brazilian Portuguese have been extensively studied topics in the literature, but new interpretations emerge through the adoption of new methodologies that would not be possible without the benefits brought about by a corpus resource segmented and aligned like C-ORAL-BRASIL. These two papers show the productivity of morphosyntactic studies that spring from data analyzed according to the parsing provided by adequate units of reference and interpreted taking into account their informational functions, which can be achieved only by integrating prosodic and textual information. In light of what has just been mentioned, it is worth reinforcing that transcriptions are actually written texts, which have as their source oral ones, and by being written do not allow the exploration of precious information that only the easy and systematic access to their audio source can provide.

Both the Italian and the Brazilian corpora from the C-ORAL family are used also in the paper by Raso & Vieira, which proposes a phonetic description of four information units within the theoretical framework of the Language into Act Theory (Cresti 2000; Moneglia & Raso 2014). This paper offers an alternative view for a much studied topic during the last three decades, i.e. the controversial category of Discourse Markers. The authors propose that prosody is the main formal feature to be employed in the identification of Discourse Markers and their functions – this contradicts the almost consensual lexicalist perspective of previous studies. Once again, this study would not be possible without perceptually segmented and text-to-speech aligned corpora, which allow the observation of how prosodic patterns convey specific functions.

The paper by Neder Neto about Brazilian EFL aspiration in voiceless stops uses the LINDSEI-BR (Mello *et al.* 2013), the Brazilian branch of the Louvain International Database of Spoken English Interlanguage[6]. The topic had been previously assessed, but never through data from spontaneous speech corpora. Therefore, once again, it is the use of spontaneous speech corpora that fosters a new methodology able to better deal with the reassessment of a previously studied topic. Together with the paper by Sarsur to be mentioned in the next section, this paper contributes to the study of second language acquisition and to phonetic teaching.

5.   Sociolinguistic and prosodic studies

Besides corpus-based speech studies, other data sources can be empirically employed for the development of the area. Lab investigations are necessary to verify findings from corpora and for experimental studies, and data collected for specific purposes are sometimes the only possible methodology. The latter is the case for the research by Silva about consonant prevocalization in coda position in Maxakali, which constitutes a specific and original case of a more general phenomenon documented in other languages. The phonological phenomenon is investigated under a sociolinguistic perspective with very interesting results which shed light on possible historical processes in the language.

Similarly to Neder Neto's paper, commented in the previous section, Sarsur's also deals with foreign language teaching, in this case French. This paper discusses how foreigners, in this case Brazilian students, can better learn the pronunciation of front rounded vowels, when these vowels are not part of their native repertoire. The paper presents an interesting methodology to improve the perception of these French phonemes in order to get better results in production.

Another interesting research in a sociolinguistic perspective is the one by Cruz & Rilliard about the role of intensity and duration in the characterization of modal intonation in the northern Brazilian Portuguese variety of the state of Pará. This research is part of the Brazilian contribution to the *Atlas Multimedia de la Prosodia del Espacio Románico* - AMPER internal project[7], which aims at comparing intonation patterns of Romance regions. The different modalities studied here are declaratives and total questions, which under alternative perspectives could be considered to be different illocutions or groups of illocutions.

---

[6] https://www.uclouvain.be/en-cecl-lindsei.html
[7] http://stel.ub.edu/labfon/amper/cast/amperespana_grupos.html

Yes/no questions are also studied here by Araújo & Rotchés, who propose a methodology for the analysis of melodic contours in Minas Gerais state Brazilian Portuguese. This is an increasingly more debated field within speech studies, since the ample variability of utterances in spontaneous speech leads to a great variability in methodologies striving to find what truly characterizes a communicative class. This can be studied as a whole, considering the differences just belonging to locutive aspects found in tokens or to characteristics of the speakers, without affecting the communicative object of study[8].

Prosody is the main object of attention also in Moura's paper on the attitude of criticism in political discourse, through the analysis of prosodic parameters. This is an example of how prosody can be studied to elucidate discourse functions under a rhetoric perspective.

In Shor's paper the relation between syntactic and prosodic parsing is discussed. The paper uses examples from the CoSIH corpus for spoken Hebrew, under a more qualitative than quantitative perspective. The paper brings to fore some of the main topics discussed in the workshop, including what the reference unit for speech should be and how it should be prosodically and segmentally defined. The main thesis of the paper lies on the importance of integrating cognitive and interactional constraints in order to explain the parsing of speech in intonation units; this means that text typology interacts with cognitive constraints in order to organize prosodic units. Therefore, the author claims the importance of observing the marked differences between monologic and dialogic prosodic structures. The segmentation criteria adopted in this paper are very similar to those adopted in the C-ORAL family corpora, described in some other papers in this collection, especially in Cavalcante & Ramos, showing interesting convergences among studies focusing on very different languages such as Hebrew on the one hand and English and Romance languages on the other.

5.   Conclusion: How this volume is organized

Multiple studies in this special issues show how speech studies have grown in the last decades in different directions, depending on the linguistic level of interest of the researchers. We would like to emphasize the growth of prosodic studies, which is presented here not only by the papers that explicitly delve into prosodic studies, but also by those that have as their goal to explain syntactic, informational or discourse aspects of speech. Nevertheless, if on one hand we can

---

[8] For different definitions of what can be called sentence modality or illocution, and different methodologies for its study in speech, see Raso & Mello (2014).

be optimistic to see more and more studies that contribute to the understanding of speech looking for grammatical constraints in all aspects of the acoustic signal, on the other hand, we still hold the impression that some linguistic categories have yet to be further discussed in their definition and scope, and that the same labels are still used to refer to different linguistic objects, while different names are used for objects which seem to be similar linguistic categories, therefore generating an overlapping of objects. We would like to highlight the necessity for discussion about the definition of the linguistic category behind the use of expressions like modality, illocution, speech act, attitude, information structure and others[9], since a messy confusion frequently originates from the use of terminology without its definition, as if it were obvious what the object behind its lexical formulation is. This is still a frequent problem in prosodic studies, which leads to difficulties in communication among different theoretical frameworks and traditions.

In this introduction we grouped the papers in a way that allowed us to make some considerations about methodologic aspects that we consider crucial for the study of speech. The organization of the volume follows a more traditional disciplinary order. The first section is dedicated to speech segmentation and corpora compilation (with the works of Cavalcante & Ramos, Mello & Mello, Mittmann & Barbosa, and Oliveira). The second one includes phonetic studies (Sarsur, Silva, Ferrari & Teixeira, Neder Neto, Raso & Vieira). The third one focuses on syntax (Bossaglia, Inbar, Silva & Mello, and Shor). The fourth and last one faces pragmatic and prosodic aspects of speech (Cruz & Rilliard, Araújo & Rotchés, Lucente, and Moura).

Many of the papers have an interdisciplinary approach, being prosody the unifying common element among them. Therefore, we invite the reader to look through the different sections, even if her/his interest lies specifically on one descriptive linguistic level.

Acknowledgments

---

[9] For a discussion of the terminology for the different categories of modality, illocution and attitude, see Mello & Raso (2011).

References

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

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. & Quirk, R. 1999. *Longman grammar of spoken and written English*. Harlow: Pearson Education.

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

Cresti, E. 2005. Enunciato e frase: teoria e verifiche empiriche. In M. Biffi, O. Calabrese & L. Salibra (eds), *Italia Linguistica: discorsi di scritto e di parlato. Scritti in onore di Giovanni Nencioni*. Siena: Prolagon, 249-260.

Cresti, E. & Moneglia, M. (eds) 2005. *C-ORAL-ROM: integrated reference corpora for spoken romance languages*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Du Bois, J.W., Chafe, W.L., Meyer, C., Thompson, S.A., Englebretson, R. & Martey, N. 2000-2005. *Santa Barbara corpus of spoken American English, Parts 1-4*. Philadelphia: Linguistic Data Consortium.

Izre'el, S. & Rahav, G. 2004. The Corpus of Spoken Israeli Hebrew (CoSIH); Phase I: The Pilot Study. In N. Oostdijk, G. Kristoffersen & G. Sampson (eds), *LREC 2004 Satellite Workshop*; Fourth International Conference on Language Resources and Evaluation: Compiling and Processing Spoken Language Corpora (Lisbon, Portugal). Paris: ELRA (European Language Resources Association), 1-7.

McEnery, T. & Hardie, A. 2012. *Corpus Linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.

Mello, H., Ávila, L., Orfanò, B.M. & Neto, T.N. 2013. LINDSEI-BR: an oral English interlanguage Corpus. In H. Mello, M. Pettorino & T. Raso (eds), *Proceedings of the VII GSCP International Conference*. Firenze: Firenze Unversity Press, 85-86.

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

Mittmann, M.M. & Raso, T. 2011. The C-Oral-Brasil Informationally Tagged Mini-Corpus. In H. Mello, A. Panunzi & T. Raso (eds), *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation*. Firenze: Firenze Unversity Press, 151-183.

Moneglia, M., & Raso, T. 2014. Notes on Language into Act Theory. In T. Raso & H. Mello (eds), *Spoken Corpora and Linguistic Studies*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 468-495.

Panunzi, A. & Mittmann, M.M. 2014. The IPIC resource and a crosslinguistic analysis of information structure in Italian and Brazilian Portuguese. In T. Raso & H. Mello (eds), *Spoken Corpora and Linguistic Studies*. Amsterdam/Philadelphia: John Benjamins, 129-151.

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

Raso, T. & Mello, H. 2014. Spoken corpora and linguistic studies: Problems and perspectives. In T. Raso & H. Mello (eds), *Spoken Corpora and Linguistic Studies*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1-24.

Raso, T., & Mittmann, M.M. 2012. As principais medidas da fala. In T. Raso & H. Mello (eds), *C-ORAL–BRASIL I: Corpus de referência do português brasileiro falado informal*. Belo Horizonte: Editora UFMG, 177-222.

Raso, T., Mittmann, M.M. & Mendes, A. C. O. 2015. O papel da pausa na segmentação prosódica de corpora de fala. *Revista de Estudos da Linguagem*, 23(3): 883-922.