

# Review

---

**Shlomo Izre’el, Heliana Mello, Alessandro Panunzi and Tommaso Raso** (eds.) (2020). *In search of basic units of spoken language: a corpus-driven approach*. (Studies in Corpus Linguistics 94.) Amsterdam & Philadelphia: John Benjamins Publishing Company. Pp. xi + 440.

**Sandrien van Ommen\***   
University of Geneva

## 1 Overview of the volume

*In search of basic units of spoken language* is a concerted effort to unite concepts and terminology in the field of corpus-based spontaneous speech segmentation and analysis. The book is a collection of papers, the result of two workshops on the definition of a single reference unit for the study of spontaneous spoken language.<sup>1</sup> The authors come from a relatively homogeneous tradition, in which they take a chiefly functionalist approach to the pragmatic nature of speech. They interface prosody with communicative content in the definition of basic units and the boundaries that encompass them. The central goal is to define the minimal stretch of speech that can have a complete communicative function; the INTONATION UNIT of Wallace Chafe (Chafe 1994), to whom the book is dedicated, can be understood to be a point of departure.

The book consists of two parts, both prefaced by an introduction. In Part I, each chapter describes a different approach to the annotation of spoken discourse, illustrated with examples for which the audio files can be consulted online. Together, the chapters cover a total of nine languages: Russian, Hebrew, Central Pomo, French, Japanese, Italian, Brazilian Portuguese, European Portuguese and German. In Part II, two American English speech fragments are segmented and analysed following the approaches described in Part I. The result of these analyses is published in an online tool, allowing for a direct comparison of the methods. The analyses are also compared qualitatively and quantitatively in the last chapter of the book.

\* E-mail: [SANDRIEN.VANOMMEN@UNIGE.CH](mailto:SANDRIEN.VANOMMEN@UNIGE.CH).

I would like to thank Anna Sara H. Romøren, Gillian Gallagher and the editors of the book for feedback on pre-final drafts of this review. Their agreement with views expressed in this final version is not implied.

<sup>1</sup> A session during the 5th International Conference on Cognitive Science in 2012 in Kaliningrad and the IX LABLITA and IV LEEL workshop in 2015 in Belo Horizonte.

## 2 Introduction

With references to the words of past and present scholars, the main introduction stresses the importance of a corpus-driven approach to spoken language, thus overcoming a 'written language bias' (Linell 2005) and '[committing] to the integrity of the data as a whole' seeing the corpus 'as more than a repository of examples to back pre-existing theories' (Tognini-Bonelli 2001, quoted on page 3). It thus makes the case for letting spoken language be central to theory formation, which entails letting go of the 'sentence' and replacing it by a different basic unit, which could for instance be the utterance, the idea unit (Kroll 1977), the intonation unit (Chafe 1994) or the conversational turn. In addition to reviewing possible levels (semantic, syntactic, prosodic, discourse) at which the basic unit can be defined, the introduction juxtaposes functionalist tradition with generative tradition to further position the book in the former, at its core differentiating the primacy of function *vs.* structure. Rather than approaching language as structured *a priori*, the authors analyse speech as consisting of communicative units whose internal structure is analysed only after their initial segmentation.

In the introduction, it is proposed that the potential risk of circularity that arises from equating the instrument and object of analysis can be circumvented with prosody. The existence of prosodic units and their availability to listeners is uncontested, as for instance evidenced by inter-rater reliability studies as well as psycholinguistic evidence of boundary perception and inference. Elemental as it might be, however, the authors consider that a basic prosodic unit is as poorly defined as it is intuitively recognised: annotators' generally high level of agreement on boundaries is not yet unambiguously or categorically captured by phonetic correlates, nor by their relation to syntax or semantics. The book aims to empirically establish common ground for the discussion of reference units of speech. To that end, it sets out to present the state of the art of research into the definition of the prosodic unit through the classification of communicative function, and considers the question of compositionality of these units in terms of the number of boundary strengths and unit functions that should be assumed.

## 3 Structure of the volume

The chapters in Part I describe methods of annotation, theories on discourse structure and quantitative approaches to discourse segmentation, always combining discourse meaning with prosodic form. Chapters 2, 4 and 6 dedicate particular space to the broader development of theory. In Chapter 2, Shlomo Izre'el explains prosody–syntax–discourse interfaces using Hebrew data, ultimately showing that the utterance encompasses each of these different layers, which do not perfectly align with each other within the utterance. In other words, the utterance is the basic unit of reference because its terminal boundary represents completion on the levels of prosody, syntax and discourse alike. In Chapter 4, Jeanne-Marie Debaisieux & Philippe Martin present the compositionality of syntactic and discourse units (microsyntax and macrosyntax respectively) and their interface with an independent hierarchically structured prosody. The approach is illustrated by the analysis of two French excerpts, and includes the description of different pitch movements and their role in compositionality. In Chapter 6, Emanuela Cresti presents examples of different illocutionary types and their prosodic realisation in Italian, first and foremost detailing a philosophical

background of language as an illocutionary act and presenting the methodology for the addition and discovery of new illocutionary types.

Chapters 1, 3, 7 and 8 focus on particular topics, languages or discourse types. In Chapter 1, Andrej A. Kibrik, Nikolay A. Korotaev and Vera I. Podlesskaya segment Russian dialogue (extending previous work on monologues) into ELEMENTARY DISCOURSE UNITS, which are argued to be the building blocks of discourse organisation. Intonational integrity drives segmentation here, as elsewhere in the book, and illocutionary types are described prosodically. In Chapter 3, Marianne Mithun segments Central Pomo using Chafe's intonation unit, developing the reasoning that the gradual nature of prosodic variables is more appropriate for the expression of the gradual nature of communicative information than it would be for that of the categorical nature of grammatical structure. In Chapter 7, the approach of Giulia Bossaglia, Heliana Mello & Tommaso Raso is based on the same theoretical background as that presented in more detail by Cresti in Chapter 6, now applied to Brazilian Portuguese. It additionally describes the combinatorial properties of intonation units in greater detail, and emphasises the pragmatic autonomy of these basic units of reference, despite their apparent syntactic dependence. In Chapter 8, Mira B. Bergelson & Mariya V. Khudyakova challenge the reliance on prosody for the segmentation of spontaneous speech. They present data from individuals with brain damage, making use of syntax for the segmentation of elementary discourse units, after which they take a descriptive approach to the data, emphasising the challenges of applying generally accepted procedures to the clinical setting.

In Chapters 5 and 9, the investigation moves to a quantitative plain. In Chapter 5, Takehiko Maruyama, Yasuhara Den & Hanae Koiso quantitatively determine a basic unit of reference for Japanese. Their analysis starts with the annotation of a preselected set of small units defined by pragmatics, prosody and syntax, among others, after which a cluster analysis reduces the number of units to SHORT and LONG UTTERANCE-UNITS, representing prosodic completion in the former case and syntactic and pragmatic completion in the latter. In Chapter 9, Plínio A. Barbosa presents an automatic segmentation procedure based on an implementation of syllable duration differences, the result of which is analysed in terms of its agreement with a human annotator. The procedure is tested on a total of five languages, including the American English excerpts that are the object of analysis of Part II, with moderate agreement for the Romance languages (approximately 0.75/0.07 hit/false alarm) and a much lower agreement for the Germanic languages.

A picture that gradually emerges is that all the authors describe the existence of similar basic units and their combinatorial properties, which in most cases entails the existence of terminal and non-terminal boundaries. They agree on the relation of prosody with these boundaries, as well as on the content of the units they delimit, where the delimiting function of prosody is defined categorically as a terminal or non-terminal contour or tone, or as integrity of the prosodic contour. Meanwhile, the units are discourse elements, and the relation of prosody to their meaning, communicative content or illocutionary force is qualitatively expressed with a set of tonal patterns. Some of the approaches encode integral melodic contours as illocutionary acts, and as such these contours can be elicited, as demonstrated by Cresti.

This supposed emergence of global agreement can be verified with the results of the annotations in Part II, in which the authors annotate the same two American English speech excerpts. These new annotations present each approach with a cross-linguistic challenge, since many of them were optimised for the description

and classification of discourse in a specific language, even if they arguably have a basis in American English as a result of Chafe's intonation unit. Part II again consists of separate chapters for separate approaches, and concludes with a comparison. This comparison, in the final chapter, quantifies agreement between the different methods in Part II as agreement on boundary placement, both overall and by pairwise comparison. To some degree, then, Part II does not just present the state of the art of discourse segmentation, but also makes an attempt at renewed convergence, at the very least with respect to the question of whether a common basic unit can be defined. Based on this comparison, the chapter concludes that terminal breaks are the most likely candidates to represent the reference unit of spoken language.

#### 4 Overall evaluation

The book is an interesting and well-conceived project, which presents different current approaches to the segmentation of spontaneous spoken language from a functionalist perspective in Part I, applying them to the same spoken fragments in Part II. The collection of papers gives an overview of a range of existing approaches to discourse segmentation, and invites renewed consideration of what constitutes a communicative act. It addresses the need for convergence of terminology, a need which becomes apparent while reading. The book reveals a dense conceptual space with a high degree of overlap, as exemplified by footnote 6 on page 45: 'Our grouping of [elementary discourse units] into illocutionary sequences, or spoken sentences, parallels (though not equals) grouping of 'utterances' into 'stanzas'; or into 'compound utterances' and further, into 'discourse patterns' (see, respectively, Cresti, this volume, Part I; Debaisieux & Martin, this volume, Part I).' References like these to the other chapters show the authors' and editors' diligence in strengthening coherence throughout the book and their commitment to the overarching goal of the project. For the reader it remains a challenge to understand to what extent and for what reason concepts are parallel but not equal, but this is arguably inevitable in an edited volume incorporating papers with different initial assumptions.

The editors signal important priorities for language research: there is still an over-reliance on written language, and prosody, for instance, is too often overlooked. Moreover, the reader is reminded that studying language as a whole and as an act of communication is central. The complexity of language invites approaching it structurally and from separate angles, but this should not come at the cost of reductionism. Meanwhile, the homogeneous theoretical embedding of the book does have an isolating effect, because it is often left to the reader to place the definitions and methodologies in the broader context of research in (if we are forced to be categorical) laboratory phonology, generative linguistics and cognitive linguistics. The introductory and concluding chapters make an attempt at comparing the approaches and placing them in context, but their mixed writing style lacks clarity. This segregation of fields is particularly problematical when it comes to the segmentation of basic units on prosodic grounds. The theoretical standpoint that function comes before form, and hence that thoughts and intentions precede structure, is well communicated. However, the delimitative role of prosody is the first step of analysis in most of the chapters, it is positioned as a crucial instrument of segmentation in the introduction, and the concluding chapter resorts to a comparison of the different methods on the basis of these segmentations, finding the terminal boundary to be the best point of

reference. The segmentation into prosodic units is therefore central in the book, but too often it is not reproducible. There is a repeated reliance on individual (and in rare cases non-native) perceptual judgement, and a lack of reference to and use of widely accepted accounts of delimitative prosody, in some cases 'proposing' well-established empirical facts, or deeming them 'likely'. Note that this does not apply across the board; e.g. Chapter 4 makes reference to autosegmental/metrical phonology (Goldsmith 1976, 1990) and gives a detailed account of its own prosodic annotation guidelines, Chapter 5 uses X-JToBI (Maekawa *et al.* 2002) and Chapter 2 gives a detailed account of phonetic boundary phenomena in the corpus at hand, although here it is unclear to what extent boundaries informed the phonetic analysis or *vice versa*. Chapter 9 has as its goal to automatically find boundaries on the basis of phonetics, a tool which can prove useful in standardisation of this task.

Overall, there are a number of well-written chapters with a clear and sufficiently contextualised theoretical outline, a reproducible analysis and interesting observations and conclusions. However, other chapters are lacking on one or more of these dimensions. This generates some uncertainty on how representative the book is as a state of the art in the field of discourse segmentation and analysis.

There are aspects of the book that may disappoint, in the sense that the overarching idea of the endeavour is so ambitious and intriguing that its implementation leaves something to be desired. The design of the endeavour, the direct comparison of segmentation approaches, is incisive and is of benefit to the field. The book contains links to the spoken fragments that are used as examples in the different chapters, which is useful and engaging.<sup>2</sup> In the case of the two American English fragments, this leads to a better understanding of the annotation choices made by the different authors. It also reveals, however, that the fragments are quite short, and for instance that the dialogue is accompanied by an activity, leading to long pauses and little ambiguity with respect to the position of boundaries. This choice of fragments poses some unfortunate limitations on the depth of comparison between the different approaches.

While some chapters in Part II can be experienced as repetitive of material in Part I, due to their descriptive nature, others make worthwhile additions. For instance, Chapter 1 in Part II adds an acoustic analysis of boundary placement by 16 annotators of different backgrounds, and Chapter 2 of Part II goes into detail on its annotation methods. The comparison of the approaches in the last chapter, finally, is where the project comes together, and where the initially stated goal is reached. The chapter compares the approaches on their segmentation focus, the relation between prosody and syntax, the theorised nature of the reference unit and finally the actual boundary placement. The reader of the entire volume looks forward to this chapter, and its qualitative comparison clarifies the nature of some inherent theoretical differences. Unfortunately, though, it remains somewhat superficial. At the same time, however, the quantitative comparison of boundary placement is emphatically thorough.

In general, the invitation to collapse or generalise over separate levels of analysis (e.g. syntax, prosody, segmental phonology, discourse, information structure) to find a single basic unit of reference can initially be surprising. The results of the book may not necessarily change that initial stance for every reader, but they do serve to offer wide-ranging views on the alignment of these levels and their relation to speech planning, and ultimately to thought.

<sup>2</sup> In some cases the direct links do not work, but all examples can be accessed via the main index.

All in all, *In search of basic units of spoken language* presents an interesting overview of approaches on the topic of discourse segmentation in different languages. It is an inspiring shared endeavour, and it contains papers which are relevant in their own right. It was conceived to address a research problem, so the book as a whole is not suited as a work of reference nor as a textbook. Researchers in the fields of corpus linguistics, discourse analysis and spontaneous speech segmentation will find papers of interest in this collection, and the quest for basic speech units will continue.

## REFERENCES

- Chafe, Wallace L. (1994). *Discourse, consciousness, and time: the flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
- Goldsmith, John A. (1976). *Autosegmental phonology*. PhD dissertation, MIT.
- Goldsmith, John A. (1990). *Autosegmental and metrical phonology*. Oxford & Cambridge, MA.: Blackwell.
- Kroll, Barbara (1977). Combining ideas in written and spoken English: a look at subordination and coordination. In Elinor Ochs & Tina Bennet-Kastor (eds.) *Discourse across time and space*. Los Angeles: Department of Linguistics, University of Southern California. 69–108.
- Linell, Per (2005). *The written language bias in linguistics: its nature, origins and transformation*. London: Routledge.
- Maekawa, Kikuo, Hideaki Kikuchi, Yosuke Igarashi & Jennifer Venditti (2002). X-JToBI: an extended J-ToBI for spontaneous speech. *Proceedings of the 7th International Conference on Spoken Language Processing (Interspeech 2002)*. 1545–1548.
- Tognini-Bonelli, Elena (2001). *Corpus linguistics at work*. Amsterdam & Philadelphia: Benjamins.