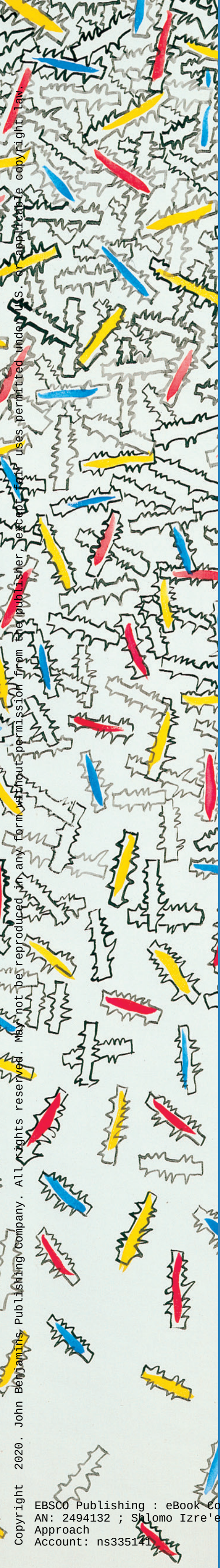


In Search of Basic Units of Spoken Language

A corpus-driven approach

EDITED BY

Shlomo Izre'el
Heliana Mello
Alessandro Panunzi
Tommaso Raso

Studies in Corpus Linguistics

94

JOHN BENJAMINS PUBLISHING COMPANY

In Search of Basic Units of Spoken Language

Studies in Corpus Linguistics (SCL)

ISSN 1388-0373

SCL focuses on the use of corpora throughout language study, the development of a quantitative approach to linguistics, the design and use of new tools for processing language texts, and the theoretical implications of a data-rich discipline.

For an overview of all books published in this series, please see *benjamins.com/catalog/scl*

Volume 94

In Search of Basic Units of Spoken Language. A corpus-driven approach
Edited by Shlomo Izre'el, Heliana Mello, Alessandro Panunzi and Tommaso Raso

In Search of Basic Units of Spoken Language

A corpus-driven approach

Edited by

Shlomo Izre'el
Tel Aviv University

Heliana Mello
Federal University of Minas Gerais

Alessandro Panunzi
University of Florence - LABLITA

Tommaso Raso
Federal University of Minas Gerais

John Benjamins Publishing Company
Amsterdam / Philadelphia

∞™ The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Cover design: Françoise Berserik
Cover illustration from original painting *Random Order*
by Lorenzo Pezzatini, Florence, 1996.

DOI 10.1075/scl.94

Cataloging-in-Publication Data available from Library of Congress:
LCCN 2019059347 (PRINT) / 2019059348 (E-BOOK)

ISBN 978 90 272 0497 4 (HB)
ISBN 978 90 272 6153 3 (E-BOOK)

John Benjamins Publishing Company · https://benjamins.com

In memory of Wallace Chafe, a master to us all

The larger goal is to achieve a better understanding of everything that makes us human, based on an awareness that language is a primary ingredient of humanness. It is exciting to realize the power of linguistics to shed light in so many ways on the richness of human experience.

(Wallace Chafe, "Searching for Meaning in Language: A Memoir," *Historiographia Linguistica* XXIX [2002]: 259)

Table of contents

Part I

Part II

▶ indicates the availability of audio files which can be found at:
https://doi.org/10.1075/scl.94.audio.

The comparative work among all the segmentations is stored in the SLAC (Spoken Language Annotation Comparison) database, through which the reader can find all the segmentations compared and analyzed, freely accessible online at <https://doi.org/10.1075/scl.94.slac>.

Acknowledgments

The editors would like to thank all participants of the LABLITA-LEEL Workshop that took place in Belo Horizonte, Brazil in August 2015, which was the seed that gave life to the long interaction among editors and authors that originated this volume. The organizing committee of the workshop was coordinated by Tommaso Raso and Heliana Mello and was supported by all the members of the Laboratory of Empirical and Experimental Linguistic Studies (LEEL) of the University of Minas Gerais and the LABLITA laboratory at the University of Florence, Italy.

The editors would also like to thank the referees of the volume, and especially Amina Mettouchi, who read and offered valuable comments regarding the the whole volume. A special thank you goes also to Marcelo Vieira, who did a great job adjusting all the papers to John Benjamins's editorial norms. We further thank Leon Shor for his substantial help with the second proofs and especially with the index. We are very grateful to Ute Römer, general editor of the series *Studies in Corpus Linguistics*, for hosting the volume in the series and for many suggestions, corrections and observations she offered. Thanks are due also to the JB staff for enabling the publication of this rather complicated volume in an orderly and successful manner, and especially and to Kees Vaes, from the editorial staff, and to Susan Hendriks, who took care of the final editorial process for their support and assistance.

INTRODUCTION

In search of a basic unit of spoken language

Segmenting speech

Shlomo Izre'el[i], Heliana Mello[ii], Alessandro Panunzi[iii]
and Tommaso Raso[ii]
[i]Tel Aviv University / [ii]Federal University of Minas Gerais /
[iii]University of Florence – LABLITA

1. Why do we need to segment speech?

This volume discusses different views about the basic unit for the analysis of speech. Notwithstanding the variation of some conceptual and applicational criteria and of terminological choice among the different groups and authors who contributed to this volume, the concept of basic unit is intended as the minimal stretch of speech that can have a complete communicative function. This is a very broad definition for what a basic unit is, and each research group or author defines better what the term means in their view in their chapters. Some authors prefer to call this entity *reference unit* or to use an even different terminology. This does not mean that these different choices point to a different entity. It is just a terminological preference for the minimal stretch of speech with a complete communication function.

This is to say that all the chapters in the volume, studying different languages in the first part of the volume and facing the same English texts in the second one, deal with a most important and basic linguistic problem: speech segmentation. This problem was neglected by linguistics for a long time, mainly because, despite manifested intentions, most linguists from nearly all theoretical schools (to a greater or lesser extent) have been conditioned by what has been called "the written language bias in linguistics" (Linell, 2005; see below).

Since the dawn of grammatical studies, written forms of language have formed the basis for looking at language (Di Benedetto, 2000, p. 395). This tradition is said to have changed during the 20th century, when linguistics started to be inclined to regard spoken language as primary, and writing as "a means of representing speech in another medium" (Lyons, 1968, p. 38).

https://doi.org/10.1075/scl.94.int

Jespersen (1924), in his preface to *The Philosophy of Grammar*, wrote:

> I am firmly convinced that many of the shortcomings of current grammatical theory are due to the fact that grammar has been chiefly studied in connection with ancient languages known only through the medium of writing, and that a correct apprehension of the essential nature of language can only be obtained when the study is based in the first place on direct observation of living speech and only secondarily on written and printed documents. In more than one sense a modern grammarian should be *novarum rerum studiosus.* (p. 7)

More than three decades later, Hockett (1958) notices:

> Old habits die hard. Long after one has learned the suitable technical vocabulary for discussing language directly, rather than via writing, one is still apt to slip. It should afford some consolation to know that it took linguistic scholarship a good many hundreds of years to make just the same transition. (p. 4)

Has the linguistic community indeed overcome tradition by the 21st century? Per Linell thinks not. In the preface to a book titled *The Written Language Bias in Linguistics*, Linell (2005) states that

> the language sciences, and in particular linguistics, have developed models and theories of language that are strongly dependent on long-time traditions of dealing with writing and written language. This … is true of present-day linguistics too, and also when spoken language is thematised. Therefore, modern linguistics is partly characterised by a paradox: there is an almost unanimous agreement on the absolute primacy of spoken language, yet language is explored from theoretical and methodological points of departure that are ultimately derived from concerns with cultivating, standardising and teaching forms of written language. (p. ix)

Still, there are growing tendencies to overcome tradition, especially when working with corpora of authentic language, be it in the written or in the spoken medium. Sinclair (2001) notes:

> 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. … [T]he language obstinately refuses to divide itself into the categories prepared in advance for it … (pp. 357–358)

Indeed, there is a big difference between two main approaches to corpus linguistics. What is implied by the quote above has been developed in relation to the so-called *corpus-driven approach*. According to Tognini-Bonelli (2001),

> the term *corpus-based* is used to refer to a methodology that avails itself of the corpus mainly to expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study (p. 65)

On the other hand, in a *corpus-driven* approach to corpus linguistics, "the linguist uses a corpus beyond the selection of examples to support linguistic argument or to validate a theoretical statement" (Tognini-Bonelli, 2001, p. 84). Explaining the corpus-driven approach further, Tognini-Bonelli (2001) says:

> In a corpus-driven approach the commitment of the linguist is to the integrity of the data as a whole, and descriptions aim to be comprehensive with respect to corpus evidence. The corpus, therefore, is seen as more than a repository of examples to back pre-existing theories or a probabilistic extension to an already well-defined system. The theoretical statements are fully consistent with, and reflect directly, the evidence provided by the corpus. (p. 84)

Of course, the study of spoken varieties needs more to release itself from the chains of tradition, and working with real data using the corpus-driven approach is essential in this respect. Similarly, Blanche-Benveniste and Jeanjean (1987) comment about tackling spoken French linguistics as follows:

> it is not a question of using spoken French to illustrate a theory, but of finding a theory that allows spoken French data to be approached. (p. 90)[1]

One of the most important methodological aspects of spoken language corpus linguistics is centered on how to segment the flow of speech. However, why is speech segmentation so important? Moreover, why the written language bias has been so strong in this respect?

If we have a linguistic sequence, say a string of words stripped off of its syntactic and semantic structure, we need to make some decisions with regard to its segmentation, if we want to know the message this sequence actually conveys. In fact, only if we segment speech can we decide what the linguistic relations within a sequence of words are. Let us look at three examples in different languages, namely English, Brazilian Portuguese and Hebrew, in order to better understand our problem.

The English sequence *people give John the book I promised him* can be segmented in many different ways, among which are the following:

(1) 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 all the sequences illustrated in (1a) to (1d), the illocutions marked between parentheses are not the only ones possible. We chose these just to give one clear sense

[1] "il ne s'agit pas d'utiliser le français parlé pour illustrer une théorie, mais de trouver une théorie qui permette d'aborder les données du français parlé" (Blanche-Benveniste & Jeanjean, 1987, p. 90).

to the sequence. Of course, in order to mark the illocution, we need other prosodic features besides those used to mark the segmentation. But our main point is that the first thing we need to do, in order to give a sense to a string of words, is to segment them in groups that have to be analyzed together and in some contrast with words grouped outside certain boundaries. By doing so, we define a first scenario, in which more interpretations are still possible, but many others are not possible anymore.

Therefore, for instance, in the English Example (1a) the seven words *Give John the book I promised him* are grouped together, while in Examples (1c) and (1d) they are distributed in two different groups. In turn, the distribution in (1c) is different from that in (1d). Example (1b) is yet another different option. This is the first reason for which Examples (1a) to (1d) cannot convey the same meaning. As already said, segmentation is not enough to decide the specific meaning of a string of words, but it is the first step, before any other can be taken. This makes speech segmentation the first operation that must be cognitively organized and communicated, using some formal features, which can be decoded by the hearer, as we shall see later.

By segmenting a string of words, we decide how many actions they convey, even if we still do not know which actions they are. Only after performing this operation, can we establish the syntactic and the semantic relations between the different words. For example, we can say that the sequence *I promised him* is a relative clause in Examples (1a) and (1b), but not in (1c) and (1d), even if we still cannot say which are the actions performed by the different segmentations. Additionally, we can analyze *People* as the subject of the verb *give* in (1b) and (1c), but not in (1a) and (1d); in turn, *give* can be analyzed as a third person plural of the indicative present form in (1b) and (1c), but must be analyzed as the second person plural imperative in (1a) and (1d). All these decisions can be taken only after the initial segmentation procedure has been achieved.

Similar considerations can be made by looking at the following examples in Brazilian Portuguese (2a–2c):

(2) 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!'

Here too, we could attribute different illocutions to each group of words. What matters, however, is that, depending on the segmentation chosen, we define the number of actions, and thus restrict the possible illocutive interpretations and establish the domain for the semantic and morphosyntactic relations among words.

João can be analyzed as subject in (2b) but not in (2a) and (2c). In (2a) and (2b) *Até amanhã* can be analyzed as an adjunct of *vai pro Rio*, but this is not the case in (2c); this changes completely the meaning of the sequence *até amanhã*, which in one case means 'until tomorrow' and in the other it is interpreted as a greeting. The verbal form *vai* must be analyzed as the second person singular imperative in (2a) and (2c), but as the third person singular of the present indicative in (2b).

Lastly, we can observe what happens regarding speech segmentation taking into account a non-Indo-European language, Hebrew, in Examples (3a) to (3f) below.

(3) a. *josef natan ve avner halχu habajta* (Assertion).
'Joseph, Nathan and Abner went home.'
b. *josef natan ve avner halχu habajta* (Question)?
'Did Joseph, Nathan and Abner go home?'
c. *josef* (Calling), *natan ve avnerhalχuhabajta* (Assertion).
'Joseph! Nathan and Abner went home.'
d. *josef* (Calling), *natan ve avnerhalχuhabajta* (Question)?
'Joseph! Did Nathan and Abner go home?'
e. *josef, natan ve avner* (Extraposed Topic) – *halχu habajta* (Assertion).
'Joseph, Nathan and Abner – they went home.'
f. *josef, natan ve avner* (Question)? *halχu habajta* (Assertion).
'Joseph, Nathan and Abner? They went home.'

These six Hebrew stretches of words, very much like as in the examples from English and Brazilian Portuguese, differ from each other due to the prosodic structure marking their different illocutions. The differences between the last two stretches (3e) and (3f) and the previous ones, namely, stretches (3a), (3b), (3c) and (3d), are also enabled by the fact that Hebrew verbs consist of full clauses, thus including also a pronominal subject. In our case, the verb *halχ-u* {went-3PL} includes a 3PL pronominal subject, so that the preceding names can function in extrapositional (topicalized or interrogative) position rather than necessarily representing the subject of the clause predicate 'went'. In the last Example (3f), the question suggests a previous mentioning of the names.

Let us further our look into speech segmentation by asking: How do we signal segmentation in speech? All the features we use for this goal pertain to what is called prosody. One of the most important functions that prosody has in language is in fact what can be called *phrasing* (Barbosa & Raso, 2018; Barth-Weingarten, 2016). Prosody is one of the basic differences between the oral and the written outputs of language. Although the study of prosody is centuries old, it has only recently found some serious hold in linguistic studies. The recognition that prosody is an integral feature of spoken language is, of course, obvious, as noted by the author of an early study of spoken English: "All words and examples are given in phonetic spelling….

Moreover, since intonation is an integral part of the grammar of Spoken English, a liberal use has been made of phonetic signs" (Palmer, 1924, p. xxxi–xxxii).

Less than two decades after that, Bloch and Trager start their syntactic analysis by looking at prosodic units:

> The analysis of constructions that involve only free forms is called SYNTAX. The first question to be answered is how we determine the limits of a construction: Where does a syntactic form begin and end?
>
> In studying a foreign language, we learn, long before we begin to make a systematic analysis of its constructions, that the utterances which are complete in themselves are of various kinds. Some are minimal free forms, words …; others are sequences of two or more free forms. We can begin their classification by taking account of the suprasegmental phonemes of juncture and intonation ….
>
> If we were analyzing the syntax of English, we should first list, from the texts which we had recorded in a phonemic notation, all the complete utterances ending in one of the four final intonations. These we should call sentences. Some sentences would be seen to contain only a single word (Go! Yes.), others more than one. Among the latter kind, there would be some with one or several non-final intonations medially; each segment of a sentence bounded by these intonations we could then define as a clause. Clauses again, we should find, may consist of a single word or of several; and the juncture between the constituent words may be open or close: the man /ðəmán/, no indeed /nôw-indíjd/….
>
> When we have delimited our syntactic units in this way, we are ready to describe their make-up in terms of the word classes (parts of speech …) which appear in them. (Bloch & Trager, 1942, p. 71)

Nevertheless, among Linell's 101points that evidence the "written language bias in linguistics", there is "the neglect of prosodies, musical dimensions and paralanguage" (Linell, 2005, point # 24). The "written language bias in linguistics" denies that prosody is an essential part of language:

> Rather, it is akin to paralanguage and, first and foremost, a property of the realization of language in *speech*. In general, linguistic signs lack a musical dimension. If, however, parts of prosodies – with a less exclusive definition of phonology – *are* to be included in the grammatical model of the language, they belong to phonology rather than syntax. (Linell, 2005, p. 60)

In the past few decades, there has been growing interest in prosody. Moreover, the growing interest in corpus linguistics and the significant developments in speech technology have enhanced endeavors to better understand the interrelationship between segmental and suprasegmental structure, between prosody and syntax, pragmatics, and discourse structure.

2. How do we segment speech?

While linguistics has concentrated its attention mainly on the function of units (i.e., groups of words that are presented together without any rupture in the acoustic signal), phonetics has paid attention to the acoustic features of such units, noting also their boundaries (see Barth-Weingarten, 2016, for an approach on the boundary from a linguistic perspective). These are two different perspectives in which segmentation can be studied: We can either focus on units as a whole or we can focus on boundaries. Of course, these different perspectives can be integrated and can enlighten each other. In this volume, the main focus is on the units, but at least two chapters (Barbosa, this volume, Part I; Raso, Barbosa, Cavalcante & Mittmann, this volume, Part II) focus on what happens at the boundaries.

Speech can be segmented at different levels, in unit types of different sizes, where each one is responsible for different aspects of speech structure. Most of the studies presented in this volume focus on the *intonation unit* (also called *prosodic units, tone unit, prosodic group* or *prosodic phrase*; see Izre'el (this volume, Part I) for terminological observations and for the term *prosodic module*); a few of them focus on different units, mainly syllables (especially those chapters interested in investigating boundaries) and stress groups (Martin). The interest in intonation unit is a corollary of the main goal of the book: to discuss what can be considered as the basic unit of speech, that is, what is the unit that conveys a minimal autonomous message and, for this reason, consists of the minimal segment of speech in the communicative sense. In this sense, the intonation unit, as we will see, plays a special role. Of course, this is not the only way in which we can segment speech. There are many important units smaller than the intonation unit, like the syllable, feet or stress groups. Moreover, we can observe not only units that are smaller than the intonation unit, but also some that are larger than it, such as complexes of units to be defined as *utterances*, or even spoken "paragraphs" and bigger discursive structures. However, it is important to consider that, whatever size of unit we are interested in, their structure always depends on the configuration of smaller units, except for the minimal ones, of course. Here we will first say something about the boundaries between intonation units, and then move to the analysis of the content and the possible functions of the intonation units themselves.

All kinds of segmentation imply the presence of a boundary, either perceived or theoretically proposed and correlated to other kinds of phenomena. In this volume, the perceptual approach predominates, but sometimes added to other theoretical assumptions. When the perceptual criterion is assumed, we should (1) show that our perception is trustable or to what extent it is trustable, and then (2) look for the physical correlates that convey this perception of boundary or break or rupture in the speech flow.

The salience of a boundary (or at least of most of them) is demonstrated through different experimental and statistic considerations. The typical test is the inter-rater agreement. When several people receive the task to segment independently the same stretch of speech, the agreement has proved to be high, usually more (sometimes much more) than 80%. In the last 15 years, several spoken corpora have been segmented into prosodic units following perceptual cues. These initiatives led to several tests for inter-rater agreement, usually measured through the Kappa test of Fleiss (1971). Some examples of spontaneous speech segmented corpora date back to the pioneering London-Lund Corpus of Spoken English (Svartvik, 1990) and reach third generation corpora like the Santa Barbara Corpus of Spoken American English (Du Bois, Chafe, Meyer, & Thompson, 2000–2005), C-ORAL-ROM (Cresti & Moneglia, 2005), C-ORAL-BRASIL (Raso & Mello, 2012), CorpAfroAs (Mettouchi & Chanard, 2010), CoSIH (<cosih.com>), C-ORAL-JAPON (Garrote et al., 2015), among others. In the case of C-ORAL-BRASIL, for example, the inter-rater agreement among three trained annotators was a Kappa of 0.86, which is an excellent score (Mello, Raso, Mittmann, Vale, & Côrtes, 2012). Inter-rater agreement may depend on the type of speech: Usually lab speech and reading are easier to segment, while spontaneous speech may lead to a lower agreement. In spontaneous speech, very interactive dialogues consisting of small turns, each with a single or just a few utterances, may prove easier to segment than long turns or monologic speech. Of course, other aspects of speech, such as speech rate or idiosyncratic features may render segmentation more difficult. However, the measured inter-rater agreement scores do not differ much and all analyzed cases confirmed the very high salience of the phenomenon considered to our perception.

A much more complex issue is to understand the physical features that convey this perception of boundaries between two intonation units. Firstly, let us try to define both *intonation unit* and its *boundary*. This is not an easy task, since the definitions we have in the literature may lead to circularity. In fact, *intonation unit* is usually defined with respect to perception of boundaries (the speech flow between two boundaries) or with respect to the coherence of a prosodic contour (Du Bois, Cumming, Schuetze-Coburn, & Paolino, 1992). On the other hand, the boundaries are defined with respect to the units they split and no clear definition is given about what a coherent prosodic contour is. Therefore, we are dealing with something (the prosodic unit with its perceptual boundary) whose existence is generally recognized, something human judgment shows strong agreement when tested, but one which is still poorly defined.

Since the precursors of contemporary prosodic research, *phrasing* called scholars' attention (Bolinger, 1965; Bloch & Trager, 1942; Lieberman, 1960; Pike, 1945). Many interesting intuitions were formulated, without the possibility of being developed and demonstrated because of lack of adequate technology. In fact, since speech is a process, and not a product like written material, we need to capture it in some stable way and with an acoustic quality that is good enough to study all the information carried in the signal. Besides, we need some way to analyze its physical acoustic manifestations (f0 variation, intensity and duration). Only the technological developments of the last decades have given us the possibility to study in detail most of the acoustic phenomena that convey the linguistic functions of speech. As for prosodic boundaries, research did not have the theoretical and technological tools, and the appropriate data, to work on spontaneous speech until very recently. Nevertheless, over some recent decades, laboratory phonetics established the groundwork that eventually led to what we can do now, namely to tackle the problem of speech segmentation in natural spontaneous speech. The availability of many third generation spoken corpora with good acoustic quality, with considerable data from different speakers, in diverse situations and speech styles, and with text-to-sound alignment, is of course a *conditio sine qua non* for this goal.

Now, we do not need to begin from zero. Laboratory phonetic studies, as well as more recent studies already working on spontaneous speech, have given us a good idea about the main phenomena that are involved in the perception of boundaries in speech (Amir, Silber-Varod, & Izre'el, 2004; Barbosa & Raso, 2018; Barth-Weingarten, 2016).

The following is a partial but satisfactory list:

1. (Silent) pause, whose presence automatically seems to convey the perception of a boundary (Martin, 1973; Mo & Cole, 2010; Shriberg, Stolcke, Hakkani-Tür, & Tür, 2000; Swerts, 1997; Tseng & Chang, 2008; Tyler, 2013);
2. Lengthening of the final syllable or syllables of a unit, that is, a decreasing of speech rate during the last syllables before a boundary (Barbosa, 2008; Fon, Johnson, & Chen, 2011; Fuchs, Krivokapić, & Jannedy, 2010; Hofhuis, Gussenhoven, & Rietveld, 1995; Mo & Cole, 2010; Silber-Varod, 2011, 2013; Tyler, 2013; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992);
3. Shortening of the first syllables of a unit, that is, speech rate increases just after a boundary (Amir, Silber-Varod, & Izre'el, 2004; Tyler, 2013), correlated with phenomena of anacrusis;
4. Reset of the f0 curve (Hermes, 2006; Thorsen, 1985, 1986);
5. Abrupt change of direction of the f0 curve (Cruttenden, 1997, among others);
6. Change of intensity at the beginning of the prosodic unit (Mo, 2008; Swerts, Collier, & Terkenet, 1994; Tseng & Fu, 2005);
7. Creaky voice and perhaps other non-modal voice qualities (Dilley, Shattuck-Hufnagel, & Ostendorf, 1996; Gordon & Ladefoged, 2001; Hanson, Stevens, Kuo, Chen, & Slifkaet, 2001; Zellers & Post, 2010).

To these parameters, at least for some languages, some phenomena of segmental nature must be added. For example, for English, final stop release or glottal closure in the vicinity of final segments may serve as a cue for the existence of a boundary (Barth-Weingarten, 2016; Dilley et al., 1996; Redi & Shattuck-Hufnagel, 2001).

However, this does not mean that presently we can capture the physical nature of a perceived boundary. Many other problems need yet to be solved. Before looking at some of them, it might be useful to say something about the pause. In fact, it is not rare for a sort of identification of pause and boundary to be found in the literature, to the point that certain phonetic and phonological traditions use the expression *virtual pause* to refer to something that is not a silent pause, but another rupture of the speech flow. Pause as a unique segmentation criterion, as it is used in some corpora (Buhmann et al., 2002; Den et al., 2010; among others; see also Moneglia, 2005, p. 24), is an easy but arbitrary criterion. It is easy because an automatic tool can easily segment big corpora if it is instructed to do so when there is silence for a certain amount of time. It is arbitrary, because we need to establish a sufficient amount of time of silence that is perceptually relevant. This duration is very variable, depending on individual and contextual factors; nevertheless, Heldner (2011) established 120 ms for a pause duration to be perceivable; an important reference can be the duration of an occlusion interval of the speaker, since pause must be longer than it, but we might take in consideration also the context where the silence occurs, in order to interpret it as pause or as occlusion. Different corpora have adopted different measures for what is called pause (Fors, 2015; Heldner, 2011; Männel, Schipke, & Friederici, 2013; Kircher et al., 2004). Moreover, the term pause does not say anything about the nature of a boundary. We will see that many scholars distinguish between boundaries that convey terminality and boundaries that convey continuity. It seems that, without taking other features in consideration, the pause alone does not help in distinguishing between these two important functions (Raso, Mittmann, & Oliveira Mendes, 2015). Finally, and this is a very relevant argument, the pause is not so common in natural spontaneous speech, in which segmentation often happens without any pause, whatever duration of silence we want to consider as a pause.

Let us go back to the problems to be solved in order to understand how perception of a break in the speech flow is made possible. The first difficulty is due to the fact that, with the exception of pauses, the presence of one of the above-mentioned features, or even more than one, does not automatically lead to the perception of break. The second one is that most of these features are not categorical. This means that we do not know the threshold we need to reach in order to perceive them as a vehicle of a boundary. If we combine these two first problems, we can easily see that a huge amount of combinations of different cues with different strengths and different weights can possibly convey the perception of break. A still different problem deals with the exact position where the physical manifestation of boundaries happens. Do the physical features we have considered happen exactly at the boundary or can they occur a little earlier or a little later? This is the reason for which it is useful to consider windows to the left and to the right of a boundary when we look for possible physical cues. All these considerations lead us to yet a different question: Do we have to consider exclusively the alternative between boundary and non-boundary, or should we consider other possibilities, such as the fact that we can have boundaries of different strengths, or of different nature and different functions (Swerts et al., 1994; Teixeira, Barbosa, & Raso, 2018)?

There is a diversity of proposals in the literature that are not necessarily mutually exclusive. Some authors prefer to differentiate boundaries according to their strength (Krivokapić, 2007; de Pijper & Sanderman, 1994). They get to distinguish up to seven different strength levels (Wightman et al., 1992), but it is dubious whether human perception can distinguish so many levels, probably due to memory limits (Cowan, 1998, 2016). Other authors divide boundaries based on whether they convey a perception of conclusion (terminal boundaries) or continuity (non-terminal boundaries). In addition, among those who choose this kind of categorization, some propose that we can have different types of terminal boundaries and different types of non-terminal boundaries. Most of the segmentation systems proposed in this volume are based on the differentiation between terminal and non-terminal boundaries. The importance of this distinction is related to whether we can consider as the main reference unit of speech – above the word level – any unit or only those sequences bound by terminal boundaries. According to the latter view, non-terminal boundaries mark units that are parsed as pertaining to the same unit. Only the terminated sequences would have enough pragmatic and prosodic autonomy to be considered as basic units of speech.

In order to conclude our overview of the phonetic problems correlated with segmentation, we need to point out that the number of variables involved is too high for human perceptual analysis alone. If we consider the different features that may be responsible for the perception of a boundary, along with the different combinations among them and their different combinations of weight, we already reach an amount of variables that humans cannot conceive without the help of technology and statistics. Nevertheless, we still have to add some other causes of further variability. It seems that actual boundary realizations are language dependent. This means that each language gives more weight to certain features and less weight to others in what phrasing is concerned; this can, probably, be explained by the specific functions, other than phrasing, that one or more features perform in different languages. For instance, Mandarin, which needs to use f0 for marking lexical tone, seems to use f0 in phrasing in a different way from English, which does not present lexical tone and can therefore use f0 more freely to mark other features (Zhang, 2012; see further Ulbrich, 2006, for different strategies among three varieties of the same language, namely, German). It also seems that boundary realization changes depending on speech style (reading, formal, informal, monologic, dialogic, etc.). We still do not know much about variability among genders or ages. How do we acquire the strategy of marking boundaries? Do we immediately acquire the complete and final feature composition? Or does our preference for different compositions change along the ontogenic path? How much can these compositions vary among different individuals (Collier, de Pijper, & Sanderman, 1993)?

This vast variability can be investigated only if we have (1) a sufficiently big amount of annotated data, (2) a computational system, and (3) a good statistical methodology. In fact, this is the approach that phoneticians have been undertaking in the last years (Avanzi, Lacheret-Dujour, & Victorri, 2008; Avanzi, Simon, Goldman, & Auchlin, 2010; Bigi & Meunieur, 2018; Christodoulides, 2018; Christodoulides, Simon, & Didirková, 2018; Ni, Zhang, Liu, & Xu, 2012; Teixeira et al., 2018; Teixeira & Mittmann, 2018; inter alia).

Some very interesting news about parsing were brought to light by psycholinguistic or neurolinguistic studies (Drury, Baum, Valeriote, & Steinhauser, 2016; Glushko, Steinhauer, DePriest, & Koelsch, 2016; Hwang & Steinhauer, 2011; Nickels, Opitz, & Steinhauer, 2013; Pauker, Itzhak, Baum, & Steinhauer, 2011; Steinhauser, 2003; Steinhauer & Friederici, 2001). Using *Event-Related Potentials* (ERP), Steinhauser, Alter, and Friederici (1999) were the first to show that perceived prosodic boundaries are associated to intervals of increased amplitude in electric activity (evoked potential), named CPS (*Closure Positive Shift*). The peak of the electric activity occurs between 400 and 800 ms after a defined moment, usually located in the last stressed syllable before the boundary. The experiments took in consideration absence and presence of pause and of other parameters which are 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 for the encephalon of the hearer to react. Currently, researchers are trying to further refine the observation of human reaction to isolated parameters or to specific parameter combinations, in order to better evaluate their effects for the perception of boundaries.

It seems that segmentation (*phrasing*) is sensitive to different modality cues, both acoustic and graphic. Cues such as commas in reading seem to cause an increase of electric activity. The phenomenon also occurs for musical segmentation, but with a greater latency; to explain this latency, the hypothesis is that it is due to lack of linguistic information, made available by the structure of the segmental content. In an acquisitional perspective, it seems that CPS is encountered only after a certain age (more or less three years of age); this is explained considering that it depends on a minimal capacity for structuring, either syntactically or prosodically.

This result is compatible with data about language acquisition (Hyams & Orfitelli, 2015; Thornton, 2016; inter alia). Interestingly, CPS seems to be more evident when the boundary is less expected; this means that when the boundary is not or is minimally predictable based on information of a nature different from prosody, the electric activity shows higher peaks. These studies seem to show clearly that prosody prevails, as a vehicle for boundaries, when it is in conflict with syntactic expectations (Bögels & Torreira, 2015; Bögels, Schriefers, Vonk, Chwilla, & Kerkhofs, 2013, 2010; see also Frazier, Clifton, & Carlson, 2004).

It is important to consider that dextral individuals have a predominant temporal processing in the left hemisphere, while spectral processes activate mainly in areas of the right hemisphere (Robin, Tranel, & Damasio, 1990; Zatorre, 1997). This is confirmed by studies on impaired individuals, showing that damage in the left hemisphere leads to loss of capacity of temporal processing (Shah, Baum, & Dwivedi, 2006). In regards to the neuronal areas involved in speech perception, both temporal cortical areas and parietal ones are bilaterally activated (Hickok & Poeppel, 2000); this is true for boundary perception too, since boundaries are marked by combinations of prosodic parameters.

3. How do we use intonation units?

With regard to the functionality of the intonation unit, we need to say that this book presents works coming from a rather homogeneous tradition. This tradition has its main focus on the pragmatic consequences of the prosodic structure of speech. This means that this tradition is firstly interested in those units that seem to immediately correlate with major communicative functions (Crystal & Davy, 1975; Halliday, 1967, 1970; among others). This tradition does not deny the existence of smaller units involved in structuring larger units, as the chapters by Debaisieux and Martin and by Martin (focused on stress groups) show. It simply means that the analysis starts from communicative units (functional analysis) and is followed by the analysis of the structures that carry such perceived functions.

There is, nevertheless, another important tradition, which focuses on the identification of structural units or domains where linguistic processes take place, and, then, recognizes that these units and processes will eventually produce clear communicative functions (Nespor & Vogel, 1986, 2007; Pierrehumbert, 1980, 2000; Pike, 1945; among many others). The generative tradition is mainly related to this second approach. We can roughly say that one important difference between these two approaches resides in the departure point of the analysis (small structural domains for the second tradition vs. large communicative units for the first one) and the direction of research (from structure to function for the second tradition vs. from the communicative function to its internal structure for the first one). While the generative tradition, which presents different views within its framework (Beckmann & Pierrehumbert, 1996; Frota, 1998; Nespor & Vogel, 1986, 2007; Pierrehumbert, 1980, 2000) looks first for minimal units and for relations among them in a structural sense, functionalist approaches look for communicative functions, such as illocutions, information units and syntactic functions, and then try to understand how they are structured internally. In other words, this last approach starts from the individualization of a clear linguistic communicative function, while the generative approach prefers to first isolate minimal entities and then look at their different functions.

The generative tradition usually adopts a system of intonation annotation, ToBI (Silverman et al., 1992), also used by some non-generative scholars, although strongly criticized by Wightman (2002) after 10 years of experience of its use. The authors of this volume do not use this system. This annotation system, among many other descriptive features, postulates five levels of disjunctures (breaks, in the ToBI terminology) above the level of the lexical word, and therefore five levels of boundaries (four if we consider only the level above the phonological word). This hierarchy cannot be compared with the difference between terminal and non-terminal boundaries, even if we accept that there are distinct types of terminal and non-terminal boundaries. What distinguishes the ToBI breaks is the strength and the salience of the disjuncture. The strongest ones, numbered 3 and 4, are the breaks that respectively correlate with intermediate phrases (ip), which are delimited by a boundary tone, and intonation phrases (IP), delimited by a boundary tone and final syllabic lengthening (Pierrehumbert, 1980).

We can roughly say that terminal breaks would always be correlated with IPs, since terminality implies also a major disjuncture. But we cannot say that all IPs would correspond to the existence of terminal breaks. Many IPs can be correlated with strongly salient non-terminal breaks. Analogously, non-terminal breaks can be correlated with IPs or ips, depending on their perceptual salience. Sometimes, however, when their salience is lower or no phonetic correlate for the strong perceptual salience can be found, they can also be annotated as a disjuncture of type 2. In the ToBI system, the most important criterion of distinction between type 3 and type 2 is either the presence or absence of a clear boundary tone, or, when such a tone is present, the clear or unclear perception of the break itself. Beckman and Elam (1997) observe that they have encountered several cases in which they felt a strong disjuncture in positions without any evidence to the expected tonal events, and also several opposite cases, in which the pitch pattern at the boundary indicated an ip or IP boundary without the pre-boundary lengthening or any cue that supported the perception of a strong break. “Break index 2 was devised to mark cases of these two types of 'mismatch' between the subjective boundary strength and the intonational constituency" (Beckman & Elam, 1997, Section 3.4.).

As we have seen, both these traditions, each coming from a different approach, take into account the perception of breaks. Nevertheless, in some views taking the generative approach the existence of prosodic domains without phonetic correlates is allowed (Nespor & Vogel, 1986, 2007). For them, in order to identify a prosodic domain, what is necessary is the presence of a phonological process referred to the domain itself.

In order to map function and prosodic form of different units, the communicative approach tends to correlate perceived f0 movement (together with other prosodic phenomena) to functional goals ('t Hart, Collier, & Cohen, 1990; Xu, 2006), attributing an important weight to f0 slope. In contrast, the generative approach first looks at a sequence of abstract target points, which can be of only two types, high (H) or low (L), and whose actual phonetic realization depends on local factors. What happens between two target points is considered as mere transitions (Ladd, 2008). The first tradition is more likely to look at prosodic forms in phonetic and gradient terms, while the second one prefers phonological and categorical decisions.

There are differing opinions as regards the way segmental and suprasegmental units are to be analyzed. Most scholars seem to share the opinion that the prosodic domain and the segmental one are to be analyzed each in itself and then look at the interface between them. Others tend to hold an integrative view of these domains. Scholars (and schools of thought) further differ with regard to the primacy of either domain, whether segmental structures direct and govern prosodic structures or vice versa, that is, that prosodic structure governs segmental structures.

For Generative Linguistics, the "sentence" is a central construct. Therefore, many studies stemming from this school and its offshoots tend to discuss prosodic structure by reference to the sentence (e.g., Selkirk, 1984). For Leveltt, "[a] surface structure has no prosody, but it does contain the information required in subsequent phases for the generation of prosodic patterns" (Leveltt, 1989, p. 170). In sharp contrast to this view, Simard and Schultze-Berndt, working in construction grammar, claim that "grammatical units in spoken language cannot, in fact, be defined without reference to prosodic units" (Simard & Schultze-Berndt, 2011, p. 153).

It should be noted, that from the recipient's perspective, prosody is a *sine qua non* when trying to delimit units of spoken language (Mettouchi, Lacheret-Dujour, Silber-Varod, & Izre'el, 2007). Prosodic units encapsulate corresponding segmental units. These segmental units overlap or otherwise interface with syntactic units. The basic prosodic unit dealt with in the majority of the literature on this topic is the *intonation unit* (see above). The tonal movement of its boundary syllable(s), together with other prosodic features, determine the discursive and interactional status of the speech unit, whether major (which indicates terminality) or minor (which indicates continuity) (Chafe, 1994, Chapter 5; Du Bois et al., 1992, Section 6; Du Bois, Schuetze-Coburn, Cumming, & Paolinoet, 1993; see further references above).

Additionally, there has been an ongoing debate as regards the central segmental domain that interacts with prosody: Is it syntax (see above for the approach of generative linguistics; see further for the syntax-prosody interface below), pragmatics or information structure (Chafe, 1994; Cresti, 2000; Cresti & Moneglia, 2010; Moneglia & Raso, 2014), some or all of them together (e.g., Halliday, 2014; Steedman, 2000), aside other domains (e.g., conversational units; Barth-Weingarten, Reber, & Selting, 2010; Schegloff, 2007)?

Likewise, there are differing opinions as regards the level of units that are comparable within the segmental and prosodic domains; for example, whether a clause or a sentence is comparable to an *intonation unit*, among other configurations. The most widespread view is that *clause* and *intonation unit* are the units that interact most (Chafe, 1994; Halliday, 2014; among many others). Along the way, the usefulness of the notion of *sentence* for the analysis of spoken language has been doubted by some authorities (e.g., Halliday, 2014; Kibrik & Podlesskaya, 2008; Miller & Weinert, 1998).

Different areas of studies have dealt with the characterization of the reference unit for the analysis of spoken language, as discussed by Foster, Tonkyn, and Wigglesworth (2000), notably semantic, prosodic (in their terminology: intonational), and syntactic approaches. Another important front of studies, which may combine different disciplines to tackle spoken language, is represented by Conversational Analysis (cf. Sidnell & Stivers, 2014) and Interactional Linguistics (cf. Selting & Couper-Kuhlen, 2001).

In a review of the assignment of units for the study of speech, Foster et al. (2000) report their consulting 87 studies, through which they concluded that the proposals available can be grouped into three major fields, namely, semantics, prosody and syntax. The authors propose their own unit, which falls within a syntactic view of the organization of speech. The groupings identified are not thoroughly discussed, however, leaving doubts as to their internal consistency. As a way to have an approximate characterization of possible units, the listing proposed by the authors are here reproduced.

Semantic approaches to the reference unit of spoken discourse focus on the identification of meaning chunks, following different parameters. The major units identified are: (a) proposition: a unit containing at least one argument and its predicate (Sato, 1988, p. 375); (b) C-unit: utterances of any kind which provide referential or pragmatic semantic meaning (Pica, Holliday, Lewis, & Morgenthaler, 1989, p. 72); (c) Idea unit: chunk of information corresponding to a cognitive or psychological reality of the speaker (Kroll, 1977, p. 85).

As for prosodic (intonational) approaches, Foster et al. (2000) compiled three main units. They are: (a) tone units/phonemic clause: configuration of pitches with a nucleus or prominence bearing syllable or syllables (Crystal & Davy, 1975, p. 16); (b) idea unit: Focusing on intonation, these units end in a clause final contour, usually followed by a pause (Chafe, 1980, pp. 13–14); (c) utterance: speech unit subjected to either an intonational contour, or bounded by pauses, or even constituting a semantic unit (Crookes & Rulon, 1985) – here Foster et al. (2000) include a criterion not compatible with their own nomenclature characterized by prosodic parameters.

Syntactic approaches cover: (a) sentences: mentioned but disregarded under the assumption that they cannot be coherently characterized in spoken data; (b) idea unit: clauses, including subordinate and relative ones (Kroll, 1977, p. 90); (c) t-unit: a main clause and its dependent clauses (Hunt, 1965, p. 20). Foster et al. (2000) present their own syntactic unit, named the Analysis of Speech Unit (AS-unit). An AS unit "is a single speaker's utterance consisting of *an independent clause, or sub-clausal unit*, together with any *subordinate clause(s)* associated with either" (Foster et al., 2000, p. 365; italics in the original).

Beyond the proposals for the study of spoken language from a semantic, syntactic and prosodic (intonational) point of view, in isolation or in combination, the study of talk as a social activity that demands some order, ensued the necessity to posit a unit of analysis also in this domain.

Conversation Analysis (CA), a field that is anchored in the study of talk in human interaction, adopts speakers' turns as their basic analytical unit. Turn-taking in an interaction, be it orderly, or interrupted or overlapping, provides the analytical domain for CA. Sidnell (2011) points out that CA

> is a set of *methods* for working with audio and video recordings of talk and social interaction. These methods were worked out in some of the earliest conversation-analytic studies and have remained remarkably consistent over the last 40 years. (Sidnell, 2011, p. 20)

To be fair to CA, its practitioners' goals are associated to the uncovering of "how participants understand and respond to one another in their turns at talk, with a central focus on how sequences of action are generated" (Hutchby & Wooffitt, 1998, p. 14). Therefore, CA does not have as its goals to explain the organization of spoken language from a strictly linguistic point of view. Its purposes fall within the sociology of verbal interaction, aided by linguistic analysis as one of its several methods.

Another framework dedicated to the study of spoken data as a means of human social activity is Interactional Linguistics. This field is interdisciplinary and greatly profits from CA constructs, most notably, the turn as a departing analytical point. According to Lindström (2009):

> Interactional linguistics builds on the same assumption as CA, namely, that ordinary conversation is an ordered, structurally organized phenomenon, and that the structures of language on different levels are subordinated, molded or influenced by the general normative aspects of social interaction. (p. 96)

In order to conduct their research, interactional linguists work with turns and their linguistic characterization, taking into account syntax, lexis and prosody.

The debate about what should be deemed as a unit of analysis in spoken discourse as discussed in the previous sections, seems to suffer from two problems pointed out by Foster et al. (2000) in their survey, which according to the authors render comparisons and replications of studies unachievable: definitions and applications. By the former, the authors mean "ostensibly identical units are either defined in different ways, or not defined at all, or defined in a way which is too simple to be used with real spoken data" (Foster et al., 2000, p. 357). The latter refers to: "if exemplified at all, definitions are accompanied by one or two citation examples which bear little resemblance to the messy reality of speech transcripts" (Foster et al., 2000, p. 357). The points raised by Foster et al. (2000) are crucial for the discussions brought forth in this volume: In discussing reference units for the analysis of speech, common ground needs to be empirically established. In this sense, all the studies to be presented are firmly grounded on corpus data, explicitly defined analytical constructs and empirically supported methodologies.

4. The content of this book

This book aims at presenting the state of the art in the research into the aforementioned issues. Its initial seeds were planted in a session organized by Vera I. Podlesskaya during the Fifth International Conference on Cognitive Science in Kaliningrad, Russia, in 2012, entitled "Spoken Discourse Corpora as a Window on Cognitive Mechanisms of Speech Production". By and large, the session dealt with issues of segmentation of spoken discourse, and was concluded with a round table: "Theory and Practice of Spoken Discourse Segmentation".

In 2015, another meeting was convened, this time as the IX LABLITA and IV LEEL International Workshop at the Federal University of Minas Gerais in Belo Horizonte, Brazil, organized by Emanuela Cresti, Heliana Mello, Massimo Moneglia and Tommaso Raso. On that occasion, the topic discussed was "Units of Reference for Spontaneous Speech Analysis and their Correlation across Languages". Concentrating on spontaneous language stems from the awareness to the fact that spontaneous varieties may differ considerably in their structure from read or other non-spontaneous linguistic output, and from the recognition that it is spontaneous language that instruct most of human cognition.

The results of that meeting make the bulk of this book. The workshop participants all work with spoken corpora and hold to the methodology of corpus-driven research. Furthermore, all participants share a strong conviction that prosodic structure is essential for the study of spoken discourse, and each has brought into the discussion their own experience in practice and theory of their respective research language(s). The languages analyzed are: Russian, Hebrew, Central Pomo (an indigenous language from California), French, Japanese, Italian, and Brazilian Portuguese. There are also speech segmentation analyses of European Portuguese, German and, finally, English.

Most of the studies presented herewith try to achieve a general and comprehensive look at the interface between prosodic units and segmental ones, discussing matches and (apparent) mismatches between segmental and prosodic units and suggesting the best practice to look for a single reference unit for the study of spontaneous spoken language. These studies are supported by a study on insubordination, an issue strongly related to the interrelationship between syntax, pragmatics and prosodic units; a study on narrative segmentation in clinical linguistics, which adds to our understanding of the cognitive processes involved with segmentation; and a cross-linguistic study of automatic segmentation. These studies are presented in Part I of this book.

All studies are accompanied by sound files, a *sine qua non* for the study of spoken language in general, and for the study of prosodic structures in particular. The audio files related to each of the examples transliterated in the book are referred to by the icon ▶ and can be found in the book's website <https://doi.org/10.1075/scl.94.audio> within each of the respective chapters by clicking on the chapter's title in the Table of Contents.

In addition to bringing forth their analysis of their respective languages, participants were asked to analyze the same two chunks in English, so that their respective analytical methodologies can be compared. Although no absolute agreement has been reached on segmentation practices, methodologies, and the theoretical approaches that had guided segmentation, the results are not only informative, but seem to show certain common tendencies of segmentation and analyses, both prosodic and segmental. These analyses and a comparative study of the different segmentations and methodologies are presented in Part II of this book.

The first part of this volume presents a variety of approaches to this pursuit. The authors of the first two chapters suggest an integrative approach, where the interface between prosody, syntax and other features of speech come together so as to be referred to as the basic unit of spoken language. They differ, however, in their conclusions.

Chapter 1, by Andrej A. Kibrik, Nikolai A. Korotaev and Vera I. Podlesskaya, is titled "Russian spoken discourse: Local structure and prosody". Basing on their approach to spoken Russian monologic discourse, the authors now extend that study looking primarily at interactional multi-party discourse and placing the speech phenomena in the context of multichannel (multimodal) communication. The evidence analyzed is the Russian Pear Chats and Stories corpus. The Elementary Discourse Unit (EDU) is posited as a central building block of local discourse structure. Canonical EDUs coincide with clauses; in addition, subclausal, superclausal and paraclausal EDUs are found. A variety of prosodic phenomena are considered, in the first place the discourse accent. A discourse-semantic category of phase is used to account for relationships between EDUs and groups of EDUs.

Chapter 2, by Shlomo Izre’el, is titled “The basic unit of language and the interface between prosody, discourse and syntax: A view from spontaneous spoken Hebrew”. Looking at spoken language as an integrative whole, where prosody, syntax and discourse features interplay as to conveying information, Izre’el endeavors at finding the best methodology for its research by advocating that the best candidate to be regarded as the basic unit of spoken discourse is a larger unit than the EDU, which is, in prosodic terms, equivalent to the commonly accepted notion of intonation unit (in Izre’el’s terminology: prosodic module). This larger unit is the *utterance*. Arguments brought are mainly phonetic, phonological (prosodic), informational, and syntactic. In addition, arguments from pragmatics and conversation analysis are mentioned.

The next three chapters (Chapters 3–5) look at the interface between syntax and prosody differently. Chapter 3, by Marianne Mithun, is titled “Prosody and the organization of information in Central Pomo, a California indigenous language”. Mithun goes against some common theoretical frameworks, where it is assumed that prosodic structure is a direct reflection of syntactic structure. Instead, Mithun claims, though prosodic structure and syntactic structure often work in concert, they are distinct. Prosodic structure differs from grammatical structure in some fundamental ways. Prosody involves continua and can be more responsive to certain subtle differences in cognitive state, discourse context, and interactive goals. Grammar (morphology and syntax) can mark more distinctions, but these are categorical and conventionalized: An affix is either present or absent; one constituent either precedes or follows another. In this chapter, some prosodic structures, their functions, and their relation to grammatical structures are discussed with examples from Central Pomo.

Chapter 4, by Jeanne-Marie Debaisieux and Philippe Martin, is entitled “Syntactic and prosodic segmentation in spoken French”. This chapter presents first of all the analytical framework adopted for the syntactical study of spoken French productions. In line with the work of the Pronominal Approach (Blanche-Benveniste, 2010; Blanche-Benveniste, Deulofeu, Stefanini, & van den Eynde, 1984; Blanche-Benveniste, Mirelle Bilger, Rouget, & van den Eyndeet, 1990; Deulofeu, 2003;

Debaisieux, 2013), the framework postulates that three components are involved in the constitution of utterances: Two syntactical components, micro- and macrosyntax, and a prosodic component that interacts independently in the constitution of units. The second part presents the application of this framework to the analysis of a conversation excerpt and an excerpt from a monologue.

Chapter 5, by Takehiko Maruyama, Yasuharu Den and Hanae Koiso, is titled "Design and annotation of two-level utterance-units: From the viewpoint of Japanese". Here, the authors distinguish between the prosodic level and the syntactic level in assigning two separate respective units, called Short Utterance-Unit (SUU) and Long Utterance-Unit (LUU). SUUs are divided by acoustic and prosodic boundaries, being equivalent of Intonation Units (Chafe, 1994), which can be considered as basic units of speaker's planning. LUUs, on the other hand, are equal to Clausal Units (Biber, Johansson, Leech, Conrad, & Finegan, 1999) and are divided by major syntactic boundaries and/or communicative interactions. The authors characterize these two-level units as basic units of syntactic chunking and/or participants' interaction. Although distinguished in their basic classification, the authors show a design of both types of units, which consists of prosodic, clausal and non-clausal units.

The two following chapters (Chapters 6–7) adopt a pragmatic orientation for their view of the basic, or reference unit of spontaneous spoken discourse. Chapter 6, by Emanuela Cresti, is titled "The pragmatic analysis of speech and its illocutionary classification according to Language into Act Theory". According to the Language into Act Theory (L-AcT), developed by the author for at least the past two decades (Cresti, 2000), reference units for the analysis of speech have a pragmatic nature since they correspond with the activation of sensory-motor-schemas and lead to the performance of different speech act types. Taking direction from the tradition of John L. Austin, L-AcT assumes that the *utterance* is the counterpart of a speech act, and its main innovation is in considering the prosodic manifestations of spoken activity. In this approach, the processing of prosody is a mandatory step for the identification of both *utterance boundaries* and *illocutionary types* at the level of the reference unit, and for the identification of *intonation unit boundaries* and *information functions* inside the reference unit. This chapter illustrates the methodology used for the induction of illocutionary types from spoken corpora and details the pragmatic features which lead to the distinction of illocutionary subclasses and types which go beyond traditional parameters. In particular, the prosodic properties of the Comment play a role in pragmatic discernment, since *root* prosodic unit types correlate with specific illocutionary types. A case study is presented which covers four illocutionary types not foreseen in existing tag-sets, and which required empirical identification via corpora. The chapter also presents a granular distinction between the illocutionary types of *self-conclusion* and *assertion* *taken for granted* (belonging to the “weak” assertive sub-class), and *ascertainment* and *evidentiality assertion* (belonging to the “strong” assertive sub-class).

Chapter 7, by Giulia Bossaglia, Heliana Mello and Tommaso Raso, is titled “Illocution as a unit of reference for spontaneous speech: An account for insubordinated adverbial clauses in Brazilian Portuguese”. In this chapter, the authors propose a synchronic, corpus-based account of insubordination, through the analysis of adverbial clauses in Brazilian Portuguese spontaneous speech at the syntax/prosody interface. The authors insist on the crucial function of segmentation prosodic cues for linguistic analysis and, specifically, for syntactic relations. Besides, it is through prosody that illocutionary and informational values are conveyed in speech. The authors claim that insubordination can be studied without assuming the existence of a grammaticalization path or main clause ellipsis processes, given that through specific illocutionary prosodic profiles, syntactically dependent clauses are assigned pragmatic autonomy, since their structure and function must be analyzed at the illocutionary level through pragmatically autonomous prosodic contours.

The following chapter (Chapter 8) approaches the issue of a reference unit for spoken discourse from yet another viewpoint, which adds a further cognitive aspect to this investigation. Chapter 8, by Mira B. Bergelson and Mariya V. Khudyakova, is titled “Narrative discourse segmentation in clinical linguistics”. This chapter deals with segmentation, definition of basic units and annotation of the first corpus of Russian narratives by individuals with brain damage – people with aphasia and right hemisphere damage – and neurologically healthy speakers. The authors show that such parameters as pause length and intonation contours cannot be used for segmentation of impaired speech. Instead, they use syntactic criteria for identification of the basic, or – as they are called in this chapter – Elementary Discourse Units (EDUs; cf. Kibrik, Korotaev, & Podlesskaya, this volume, Part I).

The last chapter of Part I (Chapter 9), which may serve also as a transition to Part II, looks at segmentation of spoken language from a different perspective altogether. Chapter 9 by Plinio A. Barbosa, is titled “Cross-linguistic comparison of automatic detection of speech breaks in read and narrated speech in four languages”. This chapter tests an algorithm for the automatic detection of speech breaks in read and narrated speech in Brazilian Portuguese (BP), European Portuguese (EP), French and German. The algorithm does not require any kind of previous transcription or linguistic analysis such as syllable or phone labeling and segmentation, but the audio file only. It operates in two stages: The first detects vowel onsets, and the second normalizes V-to-V duration intervals for obtaining smoothed duration z-scores. The peaks of smoothed duration z-scores higher than 2.5 were considered as speech breaks. Compared against human segmentation, proportion of hits for reading (circa 70%) was better than for narration (circa 60%). As for results across languages, EP and French have a higher proportion of hits than the other two languages. A test with the English *Navy* audio file (see Part II) was also made, which revealed a hit proportion similar to German. This chapter shows that syllabic duration, though not sufficient, is a very important cue to convey perception of prosodic boundaries. At the same time, it shows that the same cue may function in a similar way in some languages, but also in a clear different way in others.

The second part of the book presents a general task performed by all the authors or teams: the segmentation and analysis of part of the same two English texts extracted from the Santa Barbara Corpus of Spoken American English (Du Bois et al., 2000–2005). These texts were chosen because their acoustic quality was good, as well as, most importantly, because one of them (*Hearts*) was a more interactive, dialogic text, while the other (*Navy*) was a more monologic one. In fact, interactive and monologic texts lead to different segmentation problems and give rise to different organizations of units, allowing a more complete presentation of the problems involved in speech segmentation and analysis. Interactive texts are characterized by a continuous variation of turns, usually made up of small complete units inside them. On the other hand, monologic speech is characterized by few or no turn changes, and inside the turn it is easy to find longer units. Also, while in the interactive exchange units always carry a strong actional force, in monologues this force is somehow weakened and syntax seems to emerge more prominently (Cresti, 2005; Raso & Mittmann, 2012).

At the beginning of the second part, some instructions are given about how to read the chapters and look for convergence and divergence. Also, the texts of the two English excerpts are provided. The comparative work among all the segmentations gave rise to the SLAC (*Spoken Language Annotation Comparison*) database,[2] through which the reader can find all the segmentations compared and analyzed. Both aspects, phrasing and unit analysis, are explored from a comparative point of view, in order to better understand the level of agreement and the reasons for the disagreements.

The final chapter by Panunzi, Gregori and Rocha shows, compares and analyzes the results of this common task from different perspectives. This comparison is partly a quantitative one and partly a qualitative one, taking into account both the respective theoretical models and the respective annotations. The reader always needs to keep in mind that the segmentation task performed in the different chapters cannot be considered exactly the same. On one side, they have in common the texts to be segmented; on the other side, each author or team uses a different theoretical perspective to perform the task. Therefore, the comparison cannot correspond to a standard inter-annotator agreement, that is, to the comparison

[2] Freely accessible online at <http://lablita.it/app/slac/>

of the level of agreement among different annotators that received the same instructions. Rather, the comparison is between different procedures to segment the same texts, in order to evaluate the common ground they partake as well as differences. For these reasons, in addition to giving a tentative measure of the general agreement using standard coefficients, the chapter focuses on the evaluation of an "inter-annotation agreement". This assessment is based on the consensus in detecting terminal and non-terminal boundaries between all the annotation teams with different annotation schemas, comparing them both altogether and pairwise. The results of the comparison show that the identification of tonal boundaries, and especially for the terminal ones, has a good level of coherence even in different annotation perspectives.

In this second part, all authors and teams apply the theoretical frameworks presented in the first part of the book to the two English texts, adding further explanations that did not fit in the original chapter (i.e., the one in Part I of this book). The majority of chapters further include an in-depth analysis of two small excerpts taken from each main text, which are useful to compare in more detail the annotation procedure followed by each team. Tables reporting the different levels of analysis adopted and the parameters considered are given in an appendix at the end of each chapter, together with the tagset and the conventions used. The chapters by Cresti and Moneglia and that by Raso, Barbosa, Cavalcante and Mittmann partake the same general framework. In this case, the differences in segmentations are more directly comparable. The chapters by Izre'el, by Maruyama, by Mithun, by Cresti and Moneglia and that by Kibrik, Korotaev and Podlesskaya, offer a better explanation of their theoretical analysis of intonation units, besides the texts segmentation. The chapters by Martin and the one by Raso, Barbosa, Cavalcante and Mittmann, besides the text segmentation, focus on two other additional aspects. Martin analyzes the stress groups inside the intonation unit, showing a different structural level that can shed light on how the structure of intonation units depends on different and smaller units of analysis (Martin, 2015, 2018). Raso, Barbosa, Cavalcante and Mittmann focus on boundaries, trying to account for their perception with the investigation of formal features that may convey this perception.

References

Amir, N., Silber-Varod, V., & Izre'el, S. (2004). Characteristics of intonation unit boundaries in spontaneous spoken Hebrew: Perception and acoustic correlates. In B. Bell & I. Marlien (Eds.), *Speech prosody 2004: Proceedings* (pp. 677–680). Nara: ISCA.

Avanzi, M., Lacheret-Dujour, A., & Victorri, B. (2008). ANALOR. A tool for semi-automatic annotation of French prosodic structure. In *ANALOR. A tool for semi-automatic annotation of French prosodic structure* (pp. 119–122). Campinas, Brazil.

Avanzi, M., Simon, A. C., Goldman, J.-P., & Auchlin. (2010). A. C – PROM: An annotated corpus for French prominence study. In *Proceedings of speech prosody 2010, prosodic prominence workshop* (pp. 11–14). Chicago, IL.

Barbosa, P. A. (2008). Prominence-and boundary-related acoustic correlations in Brazilian Portuguese read and spontaneous speech. In *Proc. speech prosody* (pp. 257–260). Campinas: RG.

Barbosa, P. A. (this volume). Cross-linguistic comparison of automatic detection of speech breaks in read and narrated speech in four languages. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Barbosa, P. A., & Raso, T. (2018). Spontaneous speech segmentation: Functional and prosodic aspects with applications for automatic segmentation. *Revista de Estudos da Linguagem*, 26(4), 1361–1396.

Barth-Weingarten, D. (2016). *Intonation units revisited. Cesura in talk-in-interaction*. Amsterdam: John Benjamins. https://doi.org/10.1075/slsi.29

Barth-Weingarten, D., Reber, E., & Selting, M. (Eds.). (2010). *Prosody in interaction*. Amsterdam: John Benjamins. https://doi.org/10.1075/sidag.23

Beckman, M., & Elam, G. (1997). *Guidelines for ToBI labelling (version 3.0)*. Columbus, OH: Ohio State University Research Foundation. Retrieved from <http://www.cs.columbia.edu/~agus/tobi/labelling_guide_v3.pdf>

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

Bigi, B., & Meunieur, C. (2018). Automatic segmentation of spontaneous speech. *Revista de Estudos da Linguagem*, 26(4), 1489–1530. https://doi.org/10.17851/2237-2083.26.4.1489-1530

Beckmann, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255–309. https://doi.org/10.1017/S095267570000066X

Blanche-Benveniste, C. (2010). *Le français: Usages de la langue parlée. Avec la collaboration de Philippe Martin pour l'étude de la prosodie*. Leuven: Peeters.

Blanche-Benveniste, C., Deulofeu, J., Stefanini, J., & van den Eynde, K. (1984). *Pronom et syntaxe. L'approche pronominale et son application au Français*. Paris: CNRS-SELAF, AELIA.

Blanche-Benveniste, C., & Jeanjean, C. (1987). *Le français parlé. Transcription et édition*. Paris: Didier Érudition, Institut national de la Langue française.

Blanche-Benveniste, C., Mirelle Bilger, M., Rouget, Ch., & van den Eynde, K. (1990). *Le français parlé: Études grammaticales.* Paris: CNRS Éditions.

Bloch, B., & Trager, G. (1942). *Outline of linguistic analysis.* Baltimore, MD: Linguistic Society of America.

Bögels, S., Schriefers, H., Vonk, W., Chwilla, D. J., & Kerkhofs, R. (2010). The interplay between prosody and syntax in sentence processing: The case of subject- and object-control verbs. *Journal of Cognitive Neuroscience*, 22(5), 1036–1053. https://doi.org/10.1162/jocn.2009.21269

Bögels, S., Schriefers, H., Vonk, W., Chwilla, D., & Kerkhofs, R. (2013). Processing consequences of superfluous and missing prosodic breaks in auditory sentence comprehension. *Neuropsychologia*, 51, 2715–2728. https://doi.org/10.1016/j.neuropsychologia.2013.09.008

Bögels, S., & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics*, 52, 46–57. https://doi.org/10.1016/j.wocn.2015.04.004

Bolinger, D. (1965). Pitch accent and sentence rhythm. In I. Abe & T. Kanekiyo (Eds.), *Forms of English: Accent, morpheme, order* (pp. 139–180). Cambridge, MA: Harvard University Press.

Buhmann, J., Caspers, J., van Heuven, V. J., Hoekstra, H., Martens, J-P., & Swerts, M. (2002). Annotation of prominent words, prosodic boundaries and segmental lengthening by non-expert transcribers in the spoken Dutch corpus. In G. Rodriguez & C. Suarez Araujo (Eds.), *Proceedings of the 3rd LREC conference* (pp. 779–785). Paris: ELRA.

Chafe, W. (1980). The deployment of consciousness in the production of a narrative. In W. Chafe (Ed.), *The pear stories: Cognitive, cultural and linguistic aspects of narrative production* (pp. 9–50). Norwood, NJ: Ablex.

Chafe, W. (1994). *Discourse, consciousness and time. The flow and displacement of conscious experience in speaking and writing*. Chicago, IL: Chicago University Press.

Christodoulides, G. (2018). Acoustic correlates of prosodic boundaries in French. A review of corpus data. *Revista de Estudos da Linguagem*, 26(4), 1531–1549.

Christodoulides, G., Simon, A. C., & Didirková, I. (2018). Perception of prosodic boundaries by naïve and expert listeners in French. Modelling and automatic annotation. In *Proceedings of the 9th speech prosody conference* (pp. 13–16). Poznań, Poland.

Collier, R., de Pijper, J. R., & Sanderman, A. (1993). Perceived prosodic boundaries and their phonetic correlates. In *Human language technology. Proceedings of a workshop held at Plainsboro, NJ* (pp. 341–345). Retrieved from <http://aclweb.org/anthology/H93-1068> https://doi.org/10.3115/1075671.1075750

Cowan, N. (1998). Visual and auditory working memory capacity. *Trends in Cognitive Sciences*, 2, 77. https://doi.org/10.1016/S1364-6613(98)01144-9

Cowan, N. (2016). Exploring the possible and necessary in working memory development. *Monographs Society Res Child*, 81, 149–158. https://doi.org/10.1111/mono.12257

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

Cresti, E. (2005). Notes on lexical strategy, structural strategies and surface clause indexes in the C-ORAL-ROM spoken corpora. In E. Cresti & M. Moneglia (Eds.), *C-ORAL-ROM: Integrated reference corpora for spoken romance languages* (pp. 209–256). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.15.08cre

Cresti, E., & Moneglia, M. (Ed.). (2005). *C-ORAL-ROM: Integrated reference corpora for spoken Romance languages*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.15

Cresti, E., & Moneglia, M. (2010). Informational patterning theory and the corpus-based description of spoken language: The compositionality issue in the topic-comment pattern. In M. Moneglia & A. Panunzi. (Eds.), *Bootstrapping information from corpora in a cross-linguistic perspective* (pp. 13–45). Firenze: Firenze University Press. Retrieved from <http://www.oapen.org/search?identifier=343705>

Crookes, G. V., & Rulon, K. (1985). *Incorporation of corrective feedback in native speaker/ non-native speaker conversation*. Technical Report No. 3, Center for second language research, Social Science Research Center Institute. Honolulu, HI: University of Hawaii.

Cruttenden, A. (1997). *Intonation*. Cambridge: CUP. https://doi.org/10.1017/CBO9781139166973

Crystal, D., & Davy, D. (1975). *Advanced conversational English*. London: Longman.

Debaisieux, J.-M. (Ed.). (2013). *Analyses linguistiques sur corpus: Subordination et insubordination en français*. Cachan: Hermès & Lavoisier.

Den, Y., Koiso, H., Maruyama, T., Maekawa, K., Takanashi, K., Enomoto, M., & Yoshida, N. (2010). Two-level annotation of utterance-units in Japanese dialogs: An empirically emerged scheme. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, & D. Tapias (Eds.), *7th international conference on language resources and evaluation (LREC 2010) proceedings* (pp. 2103–2110). Valletta, Malta.

Deulofeu, J. (2003). L'approche macrosyntaxique en syntaxe: Un nouveau modèle de rasoir d'Occam contre les notions inutiles. *Scolia*, 16, 47–62.

Di Benedetto, V. (2000). Dionysius thrax and the *tékhnē grammatikḗ*. In S. Auroux, E. F. K. Koerner, H.-J. Niederehe, & K. Versteegh (Eds.), *History of the language sciences* (Vol. 1, pp. 394–400). Berlin: de Gruyter.

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word initial vowels as a function of prosodic structure. *Journal of Phonetcs*, 24(4), 423–444. https://doi.org/10.1006/jpho.1996.0023

Drury, J. E., Baum, S. R., Valeriote, H., & Steinhauser, K. (2016). Punctuation and implicit prosody in silent reading: An ERP study investigating English garden-path sentences. *Frontiers in Psychology*, 7, 1375. https://doi.org/10.3389/fpsyg.2016.01375

Du Bois, J. W., Chafe, W. L., Meyer, C., & Thompson, S. (2000–2005). *Santa Barbara corpus of spoken American English*. Washington, DC: Linguistic Data Consortium.

Du Bois, John W., Cumming, S., Schuetze-Coburn, S., & Paolino, D. (1992). Discourse transcription. *Santa Barbara Papers in Linguistics*, 4, 1–225.

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of discourse transcription. In J. A. Edwards & M. D. Lampert (Eds.), *Talking data: Transcription and coding in discourse research*, (pp. 45–89). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76, 378–382. https://doi.org/10.1037/h0031619

Fon, J., Johnson, K., & Chen, S. (2011). Durational patterning at syntactic and discourse boundaries in Mandarin spontaneous speech. *Language and Speech*, 54(1), 5–32. https://doi.org/10.1177/0023830910372492

Fors, K. L. (2015). Production and perception of pauses in speech (Unpublished doctoral dissertation), University of Gothenburg, Sweden.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21, 354–375. https://doi.org/10.1093/applin/21.3.354

Frazier, L., Clifton, C., & Carlson, K. (2004). Don't break, or do: Prosodic boundary preferences. *Lingua*, 114, 3–27. https://doi.org/10.1016/S0024-3841(03)00044-5

Frota, S. (1998). Prosody and focus in European Portuguese (Unpublished doctoral dissertation). Universidade de Lisboa, Portugal.

Fuchs, S., Krivokapić, J., & Jannedy, S. (2010). Prosodic boundaries in German: Final lengthening in spontaneous speech. *The Journal of the Acoustical Society of America*, 127(3), 1851. https://doi.org/10.1121/1.3384378

Garrote, M., Kimura, C., Matsui, K., Moreno Sandoval, A., & Takamori, E. (2015). *C-ORAL-JAPON: Corpus of spontaneous spoken Japanese*. Berlin: De Gruyter.

Glushko, A., Steinhauer, K., DePriest, J., & Koelsch, S. (2016). Neurophysiological correlates of musical and prosodic phrasing: Shared processing mechanisms and effects of musical expertise. *PLoS ONE*, 11(5). https://doi.org/10.1371/journal.pone.0155300

Gordon, M., & Ladefoged, P. (2001). Phonation types: A cross-linguistic overview. *Journal of Phonetics*, 29, 383–406. https://doi.org/10.1006/jpho.2001.0147

Halliday, M. A. K. (1967). *Intonation and grammar in British English*. The Hague: Mouton. https://doi.org/10.1515/9783111357447

Halliday, M. A. K. (1970). Language structure and language function. In J. Lyons (Ed.), *New horizons in linguistics* (pp. 140–165). Harmondsworth: Penguin.

Halliday, M. A. K. (2014). *Halliday's introduction to functional grammar. Fourth edition revised by Christian M. I. M. Matthiessen*. London: Routledge.

Hanson, H. M., Stevens, K. N., Kuo, H-K. J., Chen, M. Y., & Slifka, J. (2001). Towards model of phonation. *Journal of Phonetics*, 29, 451–480. https://doi.org/10.1006/jpho.2001.0146

't Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual study on intonation: An experimental approach to speech melody*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511627743

Heldner, M. (2011). Detection thresholds for gaps, overlaps, and no-gap-no-overlaps. *The Journal of the Acoustical Society of America*, 130(1), 508–513. https://doi.org/10.1121/1.3598457

Hermes, D. J. (2006). Stylization of pitch contours. In S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, & J. Schliesser (Eds.), *Methods in empirical prosody research* (pp. 29–61). Berlin: De Gruyter. https://doi.org/10.1515/9783110914641.29

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4(4), 131–138. https://doi.org/10.1016/S1364-6613(00)01463-7

Hockett, C. F. (1958). *A course in modern linguistics*. New York, NY: The Macmillan Company. https://doi.org/10.1111/j.1467-1770.1958.tb00870.x

Hofhuis, E. M. F. J., Gussenhoven, C., & Rietveld, T. (1995). Final lengthening at prosodic boundaries in Dutch. In E. Elenius & P. Branderud (Eds.), *Proceedings of the XIII international congress of phonetic sciences* (Vol. 1, pp. 154–157). Stockholm.

Hunt, K. (1965). *Grammatical structures written at three grade levels*. Champain, IL: National Council of Teachers of English.

Hutchby, I., & Wooffitt, R. (1998). *Conversation analysis: Principles, practices and applications*. Cambridge: Polity Press.

Hyams, N., & Orfitelli, R. (2015). The acquisition of syntax. In H. Cairns & E. Fernandez (Eds.), *Handbook of psycholinguistics* (pp. 593–614). Malden, MA: Wiley-Blackwell.

Hwang, H., & Steinhauer, K. (2011). Phrase length matters: The interplay between implicit prosody and syntax in Korean 'garden path' sentences. *Journal of Cognitive Neuroscience*, 23(11), 3555–3575. https://doi.org/10.1162/jocn_a_00001

Izre'el, S. (this volume). The basic unit of spoken language and the interface between prosody, discourse and syntax: A view from spontaneous spoken Hebrew. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Jespersen, O. (1924). *The philosophy of grammar*. Chicago, IL: The University of Chicago Press.

Kibrik, A. A., Korotaev, N. A., & Podlesskaya, V. I. (this volume). Russian spoken discourse: Local structure and prosody. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Kibrik, A. A., & Podlesskaya, V. I. (2008). Is sentence viable? In *The third international conference on cognitive science. Abstracts* (Vol. 1, pp. 84–85). Moscow: IP RAN.

Kircher, T. T. J., Brammer, M. J., Levelt, W., Bartels, M., & McGuire, P. K. (2004). Pausing for thought: Engagement of left temporal cortex during pauses in speech. *NeuroImage*, 21, 84–90. https://doi.org/10.1016/j.neuroimage.2003.09.041

Krivokapić, J. (2007). The planning, production, and perception of prosodic structure (Unpublished doctoral dissertation). University of Southern California.

Kroll, B. (1977). Combining ideas in written and spoken English: A look at subordination and coordination. In E. O. Keenan & T. L. Bennett (Eds.), *Discourse across time and space* (pp. 69–108). Los Angeles, CA: University of Southern California.

Ladd, D. R. (2008). *Intonation phonology*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511808814

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *JASA*, 32, 451–454. https://doi.org/10.1121/1.1908095

Lindström, J. (2009). Interactional linguistics. In S. D'hondt, J.-O. Östman, & J. Verschueren (Eds.), *The pragmatics of interaction* (pp. 96–103). Amsterdam: John Benjamins. https://doi.org/10.1075/hoph.4.06lin

Linell, P. (2005). *The written language bias in linguistics: Its nature, origins and transformations*. London: Routledge. https://doi.org/10.4324/9780203342763

Lyons, J. (1968). *Introduction to theoretical linguistics*. London: Cambridge University Press. https://doi.org/10.1017/CBO9781139165570

Männel, C., Schipke, C. S., & Friederici, A. D. (2013). The role of pause as a prosodic boundary marker: Language ERP studies in German 3- and 6-year-olds. *Developmental Cognitive Neuroscience*, 5, 86–94. https://doi.org/10.1016/j.dcn.2013.01.003

Martin, P. (1973). Les problèmes de l'intonation: Recherches et applications. *Langue Française*, 19, 4–32. https://doi.org/10.3406/lfr.1973.5638

Martin, P. (2015). *The structure of spoken language. Intonation in Romance*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139566391

Martin, P. (2018). *Intonation, structure prosodique et ondes cérébrales. Introduction à l'analyse prosodique*. London: Iste Editions.

Mello, H., Raso, T., Mittmann, M., Vale, H., & Côrtes, P. (2012): Transcrição e segmentação prosódica do *corpus* C-ORAL-BRASIL: Critérios de implementação e validação. In T. Raso & H. Mello (Eds.), *C-ORAL-BRASIL I: Corpus de referência do português brasileiro falado informal* (pp. 125–174). Belo Horizonte: UFMG.

Mettouchi, A., & Chanard, C. (2010). From fieldwork to annotated corpora: The CorpAfroAs project. *Faits de Langue-Les Cahiers*, 2, 255–265.

Mettouchi, A., Lacheret-Dujour, A., Silber-Varod, A., & Izre'el, S. (2007). Only prosody? Perception of speech segmentation in Kabyle and Hebrew. *Nouveaux Cahiers de Linguistique Française*, 28, 207–218.

Miller, J., & Weinert, R. (1998). *Spontaneous Spoken Language: Syntax and Discourse*. Oxford: Oxford University Press.

Mo, Y. (2008). Duration and intensity as perceptual cues for naïve listeners' prominence and boundary perception. In *Proceedings of the 4th speech prosody conference* (pp. 739–742). Campinas.

Mo, Y., & Cole, J. (2010). Perception of prosodic boundaries in spontaneous speech with and without silent pauses. *The Journal of the Acoustical Society of America*, 127(3), 1956. https://doi.org/10.1121/1.3384972

Moneglia, M. (2005). The C-ORAL-ROM resource. In E. Cresti & M. Moneglia. *C-ORAL-ROM: Integrated reference corpora for spoken Romance languages* (pp. 1–70). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.15.03mon

Moneglia, M., & Raso, T. (2014). Notes on Language into Act Theory (L-AcT). In T. Raso & H. Mello (Eds.), *Spoken corpora and linguistic studies* (pp. 468–495). Amsterdam: John Benjamins.
https://doi.org/10.1075/scl.61.15mon

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Nespor, M., & Vogel, I. (2007). *Prosodic phonology*. Berlin: De Gruyter. https://doi.org/10.1515/9783110977790

Ni, C. J., Zhang, A. Y., Liu, W. J., & Xu, B. (2012). Automatic prosodic break detection and feature analysis. *Journal of Computer Science and Technology*, 27(6), 1184–1196. https://doi.org/10.1007/s11390-012-1295-z

Nickels, S., Opitz, B., & Steinhauer, K. (2013). ERPs show that classroom-instructed late second language learners rely on the same prosodic cues in syntactic parsing as native speakers. *Neuroscience Letters*, 557, 107–111. https://doi.org/10.1016/j.neulet.2013.10.019

Palmer, H. E. (1924). *A grammar of spoken English: On a strictly phonetic basis*. Cambridge: Heffer & Sons.

Pauker, E., Itzhak, I., Baum, S. R., & Steinhauer, K. (2011). Effects of cooperating and conflicting prosody in spoken English garden path sentences: ERP evidence for the boundary deletion hypothesis. *Journal of Cognitive Neuroscience*, 23(10), 2731–2751. https://doi.org/10.1162/jocn.2011.21610

Pica, T., Holliday, L., Lewis, N., & Morgenthaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition*, 11(1), 63–90. https://doi.org/10.1017/S027226310000783X

Pierrehumbert, J. B. (1980). The phonology and phonetics of English intonation (Unpublished doctoral dissertation). MIT.

Pierrehumbert, J. (2000). Tonal elements and their alignment. In M. Horne (Ed.), *Prosody: Theory and experiment* (pp. 11–26). Dordrecht: Kluwer. https://doi.org/10.1007/978-94-015-9413-4_2

de Pijper, J. R., & Sanderman, A. A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *Journal of the Acoustical Society of America*, 96(4), 2037–2047. https://doi.org/10.1121/1.410145

Pike, K. (1945). *The intonation of American English*. Ann Arbor, MI: University of Michigan Press.

Raso, T., Barbosa, P. A., Cavalcante, F. A., & Mittmann, M. M. (this volume). Segmentation and analysis of the two English excerpts: The Brazilian team proposal. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

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

Raso, T., & Mittmann, 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* (pp. 177–221). Belo Horizonte: UFMG.

Raso, T., Mittmann, M. M., & Oliveira Mendes, A. C. (2015). O papel da pausa na segmentação prosódica de corpora de fala. *Revista de Estudos da Linguagem*, 23, 883–922. https://doi.org/10.17851/2237-2083.23.3.883-922

Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, 29, 407–29. https://doi.org/10.1006/jpho.2001.0145

Robin, D. A., Tranel, D., & Damasio, H. (1990). Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. *Brain and Language*, 39, 539–555. https://doi.org/10.1016/0093-934X(90)90161-9

Sato, C. J. (1988). Origins of complex syntax in interlanguage development. *Studies in Second Language Acquisition*, 10(3), 371–95. https://doi.org/10.1017/S027226310000749X

Schegloff, E. A. (2007). *Sequence organization in interaction. Vol. 1: A primer in conversation analysis*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511791208

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: The MIT Press.

Selting, M., & Couper-Kuhlen, E. (2001). *Studies in interactional linguistics*. Amsterdam: John Benjamins. https://doi.org/10.1075/sidag.10

Shah, A. P., Baum, S. R., & Dwivedi, V. D. (2006). Neural substrates of linguistic prosody: Evidence from syntactic disambiguation in the productions of brain-damaged patients. *Brain and Language*, 96, 78–89. https://doi.org/10.1016/j.bandl.2005.04.005

Shriberg, E., Stolcke, A., Hakkani-Tür, D. & Tür, G. (2000). Prosody-based automatic segmentation of speech into sentences and topics. *Speech Communication*, 32(1–2), 127–154. https://doi.org/10.1016/S0167-6393(00)00028-5

Sidnell, J. (2011). *Conversation analysis: An introduction*. Malden, MA: Wiley-Blackwell.

Sidnell, J. & Stivers, T. (2014). *The handbook of conversation analysis*. Chichester: Wiley-Blackwell.

Silber-Varod, V. (2011). The SpeeCHain perspective: Prosody-syntax interface in spontaneous spoken Hebrew (Unpublished doctoral dissertation). Tel-Aviv, Israel.

Silber-Varod, V. (2013). *The SpeeCHain perspective: Form and function of prosodic boundary tones in spontaneous spoken Hebrew*. Saarbrücken: LAP Lambert Academic Publishing.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. In *ICSLP-1992* (pp. 867–870).

Simard, C., & Schultze-Berndt, E. (2011). Documentary linguistics and prosodic evidence for the syntax of spoken language. In G. Haig, C. Wegener, S. Schnell, & N. Nau (Eds.), *Documenting endangered languages: Achievements and perspectives* (pp. 151–176). Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110260021.151

Sinclair, J. (2001). Review of Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman. *International Journal of Corpus Linguistics*, 6, 339–359.

Steedman, M. (2000). *The syntactic process*. Cambridge, MA: The MIT Press.

Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and Language*, 86(1), 142–164. https://doi.org/10.1016/S0093-934X(02)00542-4

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, 2(2), 191–196. https://doi.org/10.1038/5757

Steinhauer, K., & Friederici, A. D. (2001). 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*, 30(3), 267–295. https://doi.org/10.1023/A:1010443001646

Svartvik, J. (Ed.). (1990). *The London corpus of spoken English: Description and research*. Lund: Lund University Press.

Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *The Journal of the Acoustical Society of America*, 101(1), 514–521. https://doi.org/10.1121/1.418114

Swerts, M., Collier, R., & Terken, J. (1994). Prosodic predictors of discourse finality in spontaneous monologues. *Speech Communication*, 15(1–2), 79–90. https://doi.org/10.1016/0167-6393(94)90043-4

Teixeira, B. H. F., Barbosa, P. A., & Raso, T. (2018). Automatic detection of prosodic boundaries in Brazilian Portuguese spontaneous speech. In A. Villavicencio, V. Moreira, A. Abad, H. Caseli, P. Gamallo, C. Ramisch, H. G. Oliveira, & G. H. Paetzold (Eds.). *Computational processing of the Portuguese language*, (pp. 429–437). New York, NY: Springer.

Teixeira, B. H. F., & Mittmann, M. M. (2018). Acoustic models for the automatic identification of prosodic boundaries in spontaneous speech. *Revista de Estudos da Linguagem*, 26(4), 1455–1488.

Thornton, R. (2016). Children's acquisition of syntactic knowledge. In M. Aronoff (Ed.), *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press.

Thorsen, N. G. (1985). Intonation and text in standard Danish. *Journal of the Acoustical Society of America*, 77(3), 1205–1216. https://doi.org/10.1121/1.392187

Thorsen, N. G. (1986). Sentence intonation in textual context. *Journal of the Acoustical Society of America*, 80(4), 1041–1047. https://doi.org/10.1121/1.393845

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.6

Tseng, C. Y., & Chang, C. H. (2008). Pause or no pause? Prosodic phrase boundaries revisited. *Tsinghua Science and Technology*, 13(4), 500–509. https://doi.org/10.1016/S1007-0214(08)70080-4

Tseng, C. Y., & Fu, B. L. (2005). Duration, intensity and pause predictions in relation to prosody organization. In *Proceedings interspeech 2005* (pp. 1405–1408). Lisbon, Portugal. Retrieved from <http://www.ling.sinica.edu.tw/eip/FILES/publish/2007.4.12.99500673.0143164.pdf>

Tyler, J. (2013). Prosodic correlates of discourse boundaries and hierarchy in discourse production. *Lingua*, 133, 101–126. https://doi.org/10.1016/j.lingua.2013.04.005

Ulbrich, C. (2006). Prosodic phrasing in three German standard varieties. In *Proceedings of the 29th annual Penn linguistics colloquium*, (pp. 361–373). Philadelphia, PA.

Whightman, C. W. (2002). ToBI or not ToBI. In *Speech prosody 2002, Aix-en-Provence* (pp. 25–29). Aix-en-Provence, France.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, 91(3), 1707–1717. https://doi.org/10.1121/1.402450

Xu, Y. (2006). Principles of tone research. In *Proceedings of international symposium on tonal aspects of languages* (pp. 3–13). La Rochelle, France.

Zatorre, R. J. (1997). Cerebral correlates of human auditory processing: Perception of speech and musical sounds. In J. Syka (Ed.), *Acoustical signal processing in the central auditory system* (pp. 453–468). New York, NY: Plenum Press. https://doi.org/10.1007/978-1-4419-8712-9_42

Zellers, M., & Post, B. (2010). Aperiodicity at topic structure boundaries. *Proceedings of speech prosody 2010: Fifth conference* (pp. 451–480). Chicago, IL.

Zhang, X. (2012). A comparison of cue-weighting in the perception of prosodic phrase boundaries in English and Chinese (Unpublished doctoral dissertation). University of Michigan.

Part I

CHAPTER 1

Russian spoken discourse

Local structure and prosody

Andrej A. Kibrik[i], Nikolay A. Korotaev[ii] and Vera I. Podlesskaya[ii]
[i]Institute of Linguistics RAS and Lomonosov Moscow State University /
[ii]Russian State University for the Humanities

We previously developed an approach to spoken Russian monologic discourse, and are now extending that, looking primarily at interactional multi-party discourse, contextualizing speech phenomena as multichannel (multimodal) communication. The evidence analyzed is the Russian Pear Chats and Stories corpus, see <multidiscourse.ru>. Scores transcripts are introduced to annotate the interlocutors' shared time line, including periods of silence. The elementary discourse unit (EDU) is posited as a central building block of local discourse structure. Canonical EDUs coincide with clauses; additionally, subclausal, superclausal, and paraclausal EDUs are found. Prosodic phenomena are considered; EDUs and groups of EDUs are accounted through a discourse-semantic category of phase. Disfluencies and other structural phenomena are systematically treated. Conventions of discourse transcription capture both prosodic and functional aspects of discourse.

Keywords: spoken discourse, discourse transcription, local discourse structure, elementary discourse unit, prosody, pause, discourse accent, phase, spoken sentence

1. Introduction

In the course of our previous work, we developed an approach to Russian spoken discourse, having introduced the notion of elementary discourse unit (EDU) and paying attention to a variety of prosodic phenomena. That approach was based on monologic, audiorecorded speech. The main result of that previous work is the book "Night Dream Stories: A Corpus Study of Spoken Russian Discourse" (Kibrik & Podlesskaya, 2009; in Russian).[1] The Night Dream Stories corpus, along with

[1] The Night Dream Stories corpus is made up of Russian spoken stories, told by children and adolescents about their night dreams. The corpus contains 129 stories, which are presented in audio and transcribed formats.

https://doi.org/10.1075/scl.94.01kib
Additional files available at http://doi.org/10.1075/scl.94.01kib/audio

several other Russian corpora, are available at <spokencorpora.ru>. See also partial English accounts by Kibrik (2011) and Podlesskaya (2011a).

In this study we extend our earlier approach to unrestricted multi-party, interactive discourse. Logically, monologue is not opposed to dialogue, it is a subtype of dialogue in which turn-taking does not happen or is very limited. Therefore, we do not believe that different principles of understanding (or transcribing) interactive multi-party discourse and monologue need to be implemented. Rather, if one proposes a comprehensive approach to interactive discourse, the corresponding principles and methods would apply to monologue, with appropriate simplification.

The analysis reported here constitutes a part of a larger enterprise we are currently involved in, namely a study of multichannel (multimodal) Russian discourse (see Kibrik & Fedorova, 2018; Fedorova & Kibrik, 2020). In that project we are creating a resource, consisting of audio, video, and eyetracking recordings of natural multichannel communication, along with annotations of various kinds of behavior: verbal, prosodic, oculomotor, hands and head gesticulation, among others. Thus, our present understanding of talk is embedded in the broader context of multichannel communication.

The resource under construction is composed of partly structured communicative exchanges, associated with the well-known stimulus material: the Pear Film (Chafe, 1980; Kibrik, 2015).[2] This six-minute film produced by Chafe and his colleagues at the University of California at Berkeley in the 1970s was originally intended to elicit stories from speakers of various languages. The film was constructed so that its scenes incline participants to describe agentive and non-agentive events, explain cause-effect relations, and account for the characters' thoughts and emotions. In accordance with the developed design of data collection in our current project, the sessions of communicative exchanges were organized as follows. Each session involved four participants with fixed roles: the Narrator, the Commentator, the Reteller, and the Listener. At the very beginning the Narrator and the Commentator each watched the film on a personal computer, trying to memorize the plot as precisely as possible. Then the main stages began. First, the Narrator told the Reteller about the plot of the film; this is a monologic stage – *first telling*. During the subsequent, interactive, stage – *conversation* – the Commentator added details and corrected the Narrator's story where necessary, and the Reteller checked his/her understanding of the plot, asking questions to both interlocutors. Then the Listener joined the group and another monologic stage – *retelling* – followed, during which the Reteller was retelling the plot of the film to the Listener. Finally, the Listener wrote down the content of the film. Forty sessions have been recorded between 2015 and 2017, and multichannel annotation is gradually evolving into

[2] <www.linguistics.ucsb.edu/faculty/chafe/pearfilm.htm>

the resource titled "Russian Pear Chats and Stories" (RUPEX). Current results can be seen at <multidiscourse.ru>.

The analysis in this chapter is based on three sessions already annotated in RUPEX, as well as on the data of our prior work. The chapter is structured as follows. In Section 2, we discuss the general organization of the vocal component of discourse, including the contrast between vocalization and silence, sequencing of turns, and the subdivision of the vocal signal into the verbal and the prosodic channels. Section 3 explains the notion of elementary discourse unit, fundamental to our approach. Sections 4 to 7 are devoted to the most central kinds of structural and prosodic phenomena found in spoken discourse. Some other phenomena are mentioned in Section 8. Since we are currently interested in how speech interacts with other types of multichannel communicative behavior, we address such interaction at certain points when discussing speech phenomena. We return to this issue in a more general way in the concluding Section 9 and provide further details on the relationship between the vocal and the kinetic communication channels. At the end of the chapter two appendices are found: Appendix A, a list of transcription conventions; and Appendix B, a transcript of an interactive excerpt, 42 sec long, from one of the sessions of Russian Pear Chats and Stories.

2. General organization of vocal discourse

When we talk, we keep silent during a significant part of the time (Goldman-Eisler, 1972; Jaworski, 1997; Krivnova, 2007; Tannen & Savile-Troike, 1985; inter alia). Periods of *vocalization* alternate with periods of silence, or pauses. *Pauses* are useful to a speaker as they are periods of time to both inhale and plan a subsequent portion of talk. As is discussed below (Section 3.1), pauses are among the criteria helping to chunk talk into elementary discourse units (EDUs), that is elementary steps in discourse production. There is a well-grounded tradition of attributing boundary pauses to subsequent EDUs (Chafe, 1994; Du Bois, Schuetze-Coburn, Cumming, & Paolino, 1992; Kibrik & Podlesskaya, 2009): It is during this time that a speaker formulates the cognitive plan for producing a subsequent EDU. However, in multi-party discourse, boundary pauses cannot be attributed to particular EDUs and even to particular speakers; see the practice adopted in Conversation Analysis to annotate pauses in separate lines (e.g., see Jefferson, 2004). Consider a moment of speaker alternation: X was the speaker before, and Y starts speaking now. If there is a pause between the vocalizations of X and Y, this pause is impossible to interpret specifically as belonging to Y's first EDU. Y could have started planning his/her EDU well before the end of X's talk, or sometime after X is done with his/her contribution. Therefore, we have to posit the idea of shared silence. In the conversational data we have analyzed so far, shared silence occupies 11.3% of the time.

In transcripts such as in Appendix B, there are two columns for marking shared pauses. The first of these contains a pause's ID-number, and the second one indicates its duration (in seconds). (For users' convenience, the representation such as in Appendix B also contains the leftmost column with numbers of specific graphic lines; numbering goes from the beginning of the whole session transcript.) In multi-party discourse, interlocutors also have a shared timeline, marked in two further columns in transcripts. These two columns contain data on the beginnings and ends of all relevant intervals, including pauses (to the left) and vocalizations (to the right). These numbers indicate times from the beginning of the sound file. We measure time with the precision of 10 ms, in accordance with the general principles developed for the RUPEX project (Kibrik & Fedorova, 2018).

In principle, it is technically possible to organize transcripts of multi-party discourse in separate sheets, one for each interlocutor. Below we use this format in those examples that are fully monologic. However, generally we find it more illuminating to use *scores* transcripts, such as the one shown in Appendix B. This kind of representation demonstrates readily the dynamics of vocal events as they unfold in time. In the transcript, there are two columns per each interlocutor. (As was mentioned above, there are three main interlocutors: the Narrator (N), the Commentator (C), and the Reteller (R).) The first column marks an ID-number of the given EDU, while the other (main) column contains the verbal and prosodic contents of the EDU.[3]

Figure 1 depicts a possible ideal sequence of pauses and vocalizations, in the case of two interlocutors.

In this kind of ideal situation, the following two things hold:

1. Interlocutors never speak at the same time;
2. Periods of silence and vocalization always strictly alternate.

However, both of these tenets are regularly violated in real life. For example, consider EDUs R-vE021 and C-vE020 in Appendix B (lines 0496 to 0499). The latter starts 890 ms before the former ends. This is an instance of overlap in the talk of two interlocutors. In the three conversations analyzed so far, overlaps take 15.6% of the overall time, including 1.3% of the time when all three interlocutors talk (cf. lines 0522 to 0523 and 0544 to 0545 in Appendix B).

When there are several consecutive EDUs produced by one and the same speaker, these EDUs may or may not be interspersed by boundary pauses. Consider

[3] An EDU's ID-number consists of several sections. First, a capital letter (N, C, or R) indicates the role of the participant. Second, the "-vE" section stands for *vocal channel EDU*. Finally, the number of the EDU is provided in the 000 format. Other phenomena, such as isolated filled pauses, laughs and other non-verbal vocal events, have separate numbering.

Pauses	Interlocutor A	Interlocutor B
Pause		
	EDU A1	
Pause		
	EDU A2	
Pause		
	EDU A3	
Pause		
		EDU B1
Pause		
		EDU B2
Pause		
	EDU A4	
Pause		

Figure 1. Ideal delivery in multi-party discourse

the sequence C-vE024 to C-vE028 (lines 0510 to 0525 in Appendix B). The first three EDUs of this sequence are pronounced without boundary pauses, apparently in a single breath. However, further EDUs are separated by boundary pauses (vp096 and vp097; lines 0517 to 0519). In transcripts such as in Appendix B, intervals of continuous vocalization, comprising no boundary pauses, are shown with color filling, differentiating the interlocutors.

The vocal signal consists of contributions belonging to two vocal channels: verbal and prosodic. The verbal contribution boils down to a sequence of phonetic realizations of phonemes. In our work, we use the standard Russian orthography to convey the verbal structure. Russian orthography is represented in strict Roman transliteration in this chapter. In Appendix B, English translation is also provided. (Examples within the main text of the chapter, in addition, include simplified word-by-word glosses.)

Prosody is a cover term for non-verbal aspects of sound. A variety of prosodic phenomena, as well as corresponding notation conventions, are discussed below in Sections 3 to 8.

3. Elementary discourse units

3.1 Identification

It has been known at least since the middle of the 20th century (see a brief review and references in Kibrik, Korotaev, & Podlesskaya, this volume, Part II) that spoken discourse is produced in a stepwise fashion. Speech does not flow like water in a quiet river. Rather it progresses in spurts, or quanta. In this way spoken discourse is analogous to much more basic kinds of goal-oriented behavior, including in other mammals. As is discussed in Kibrik (2011, pp. 281–282), this kind of analogy reveals that the quantized nature of spoken discourse has deep evolutionary and neurophysiological roots.

Among the various terms applied in the literature to the quanta of spoken discourse, we prefer the term *elementary discourse unit* (EDU). Unlike more form-oriented terms, such as *intonation unit*, this term emphasizes the constructional role of these units in discourse organization and production.

EDUs are identified primarily on prosodic grounds (see Kibrik & Podlesskaya, 2009; Korotaev, 2015). Figure 2 demonstrates a sequence of three EDUs from the Funny Stories corpus, 31_f: E002 – E004.[4]

3.29	p001	(0.86)	
4.14	E002	Moej /dočeri bylo pjat' /let, my daughter was five years	
5.34	E003	moemu /plemjanniku bylo desjat' /let, my nephew was ten years	
6.85	p002	(0.27)	
7.12	E004	i /oni s /utra eli \kašu, and they from morning ate cereal	Creaky voice, especially near the EDU boundaries.
'My daughter was five years old, my nephew was ten years old, and in the morning they were eating cereal…'			

Figure 2. EDU exemplification (source: Funny Stories, 31_f: E002–E004)

Two of the three EDUs in Figure 2 (E002 and E004) are preceded by boundary pauses. Each EDU contains one primary discourse accent (represented in transcript by underlining the accented vowel); see Section 4 below on the direction of pitch in accents. Primary accents mark informational centers of EDUs, so-called rhemes.[5] EDUs are typically characterized by intonational integrity: The f0 contour starts at an intermediate level (typical of the given speaker's voice), then has one or more

[4] This example is a monologic fragment, taken from one of our earlier corpora (available at <www.spokencorpora.ru>). These kinds of examples follow a somewhat different format than conversational fragments. In the leftmost column, only the beginning of the corresponding EDU/pause is indicated. The rightmost column may contain specific comments on the given EDU. An English translation of the whole fragment is provided in a separate line that concludes every example.

[5] Regarding the complicated notion of rheme, cf. the following recent statement in Fernandes-Vest (2016): "Using this terminology we follow a long tradition of European Theme-Rheme studies promoted specially by the Prague School functionalists and further developed in typological studies of discourse organization" (pp. 10–11); also cf. Sornicola (2006).

peaks, and often descends towards the end. For example, EDU E004 in Figure 2 has the f0 contour as shown in Figure 3.

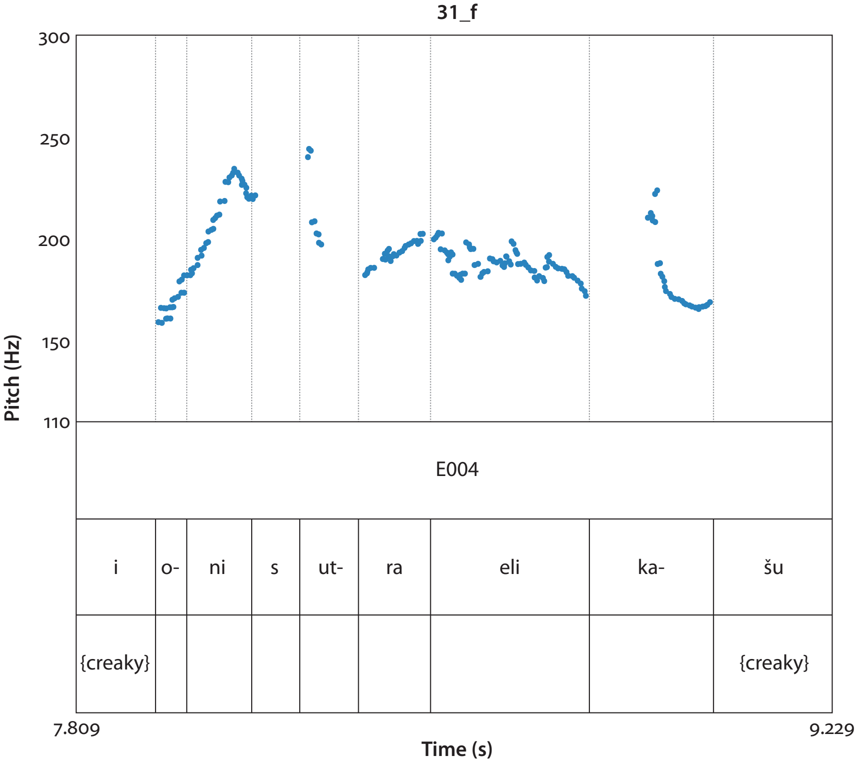


Figure 3. F0 contour of EDU E004 in Figure 2

The example in Figure 2 is also a good illustration of another highly important criterion of EDU identification: the tempo pattern. EDUs typically start with accelerated tempo and decelerate towards the end. As the data in Table 1 demonstrate, mean syllable durations in EDUs' final part are 1.5 to 2 times longer compared to the initial part.

Table 1. Tempo variation in the EDUs in Figure 2

EDU#	Initial part		Final part	
	Total, s	Per syllable, s	Total, s	Per syllable, s
E002	0.47	0.09	0.73	0.18
E003	0.66	0.09	0.85	0.17
E004	0.58	0.12	0.66	0.17

From a multichannel perspective, it is interesting to note that individual gestures are identified in the stream of kinetic behavior on the basis of segmentation principles similar to those used in EDU identification. In particular, gestures have an integral trajectory, analogous to an f0 contour, and are organized around an effort peak (so-called stroke), analogous to a primary accent.

3.2 EDUs and clauses

Prosodically identified EDUs tend to correlate with clauses (Chafe, 1994; Croft, 1995; Ford & Holmes, 1978; Kibrik & Podlesskaya, 2006; Levelt, 1989; Pawley & Syder, 2000; Thompson & Couper-Kuhlen, 2005). In Chafe's conversational corpus (Chafe, 1994), clausal EDUs (clausal intonation units, in his terms), accounted for 60% of the overall number of EDUs. In the Russian Night Dream Stories corpus, the level of correlation was even higher, about 68% (Kibrik & Podlesskaya, 2009, p. 371). Studies of various discourse genres and languages (Iwasaki & Tao, 1993; Markus, 2009; Matsumoto, 2003; Wouk, 2008) found somewhat lower or still higher levels of correlation. It is clear that language users tend to produce units of their behavior (EDUs) so that they align with the units of memorized experience (that is, clauses). See also Izre'el (this volume, Part I) for a detailed discussion of such correlation in Hebrew and in a cross-linguistic perspective.

The example in Figure 4 starts with two clausal EDUs. The first one is a main clause, and the second is a dependent (relative) clause. Further follow two EDUs that are smaller and larger than a clause.

287.59	N-vE172	(ˀ 0.32) /I-i (ə 0.69) dal'̌še my eščë vidim \ ↑ fermera^{ɯ}, and further we also see farmer
290.92	pN-065	(0.69)
291.61	N-vE173	(ɐ 0.11) kotoryj \ ↑ spuskaetsja^{ɯ}, who descends
293.21	pN-066	(0.38)
293.59	N-vE174	s \ ↑ lestnicy, from ladder
294.23	N-vE175	i vidit čto-o /ˀodnoj-j iz korzin ne \xvataet^{h}. and sees that one of baskets not suffice

'And then we also see the farmer, who descends the ladder and sees that one of the baskets is missing'

Figure 4. EDU-clause illustration (source: pears04: N-vE172–N-vE175)

EDU N-vE174 is *subclausal*; more specifically, it is an instance of what we call *increment*. After the speaker uttered the clause in N-vE173, she realized that she wanted to add an adjunct to the clause; the pause pN-066 apparently is the time during which this decision was made. The adjunct grammatically belongs to the base clause just uttered, but prosodically it is a separate EDU.

In contrast, EDU N-vE175 is *superclausal*: It includes a matrix clause and a complement. This biclausal construction is produced in a single prosodic complex and constitutes one EDU.

In the Night Dream Stories corpus, subclausal and superclausal categories accounted for 26% and 6% of all EDUs, respectively; see Kibrik & Podlesskaya (2009, p. 371) and Kibrik (2011) for further details. More recently (Podlesskaya, 2011b) we have somewhat reconsidered our approach. First, in the system we are using now the share of superclausal EDUs becomes somewhat greater; in particular, in Kibrik & Podlesskaya (2009), instances such as N-vE175 in Figure 4 were treated as a pair of monoclausal EDUs, the first one lacking a primary accent (see also Bossaglia, Mello, & Raso, this volume, on a sophisticated interplay between syntactic and prosodic segmentation in multiclausal sequences). Second, now a subset of former subclausal EDUs are considered *paraclausal*: semantically and syntactically deficient EDUs (cf. Švedova et al., 1980, para. 2674–2679; Yanko, 2008), including formulaic utterances, holophrases, interjections, vocatives, onomatopoeias, among others. Some examples can be seen in Appendix B: EDUs N-vE224, R-vE026, R-vE023, C-vE032, or R-vE020.

4. Accents, pitch, and phase

Some words bear *discourse accents*. The crucial overt manifestation of an accent is prosodic prominence: a relatively strong expiratory pulse. This pulse is usually associated with the lexically stressed syllable of a word in question. In Russian, accents are often conveyed not just by pulse alone, but also by *pitch* movement on the given syllable. Accents and pitch movements are indicated in our transcription system by means of unified symbols: / for rising pitch accent, \ for falling pitch accent, and – for level pitch accent. More complex pitch configurations, such as \/ or /– also occur. Moreover, we indicate significant pitch movements before and after the accented syllable. For example, \↑ in EDU N-vE172 in Figure 4 means that the stressed first syllable of the word *fermera* bears a falling pitch accent, while the intonation curve rises on the subsequent syllables. If just \ were marked in this case, that would mean that the intonation curve keeps descending or is level after the accented syllable.

The first three EDUs in Figure 4 contain one accent each. However, the final EDU N-vE175 contains two accents. In such instances, one accent is usually the *primary* one, while the others are secondary. The primary accent is rhematic. In our transcription system, we distinguish the primary accent from the rest by underlining the word's stressed vowel; see the final EDU in Figure 4, in which the stressed syllable of *xvataet* is underlined, while nothing is underlined in *odnoj*, bearing a secondary accent. (In fact, the recognition of secondary accents may suggest prosodic groupings inside EDUs, cf. an idea of distinguishing between "short-utterance units" and "long utterance units" in Maruyama, Den, & Koiso, this volume, Part I.)

The direction of pitch in an accent is responsible for the discourse-semantic category of *phase*, introduced by Kodzasov (1996, 2009); it is close to *transitional continuity* in Du Bois et al. (1992, pp. 28–31). This category conveys abstract semantics "anticipated continuation versus completion" (of something). Phase can be observed at three different hierarchical levels of discourse constituents. First, and most broadly, "anticipated continuation" (and the corresponding rising pitch) may refer to a speaker's illocutionary act projecting a continuation. The most typical illocutionary act of this sort is a yes/no question. Conversely, the abstract semantics of *completion* (and the corresponding falling pitch) may refer to an illocutionary act that does not project a necessary continuation, in particular a statement; see Section 5 below. Second, the direction of pitch in a primary accent may be due to an EDU's role within an illocutionary chain; see Section 6 below. Third, the direction of pitch may have the most narrow function: a relationship between EDU constituents such as theme and rheme. For example, in the final EDU of the example in Figure 4 the rising pitch accent on *odnoj* anticipates a rhematic conclusion towards the end of the structure. Both of the accents in N-vE175 can be seen in Figure 5. The rather high rise in the thematic accent on *odnoj* is known in the Russian tradition of intonology initiated by Elena A. Bryzgunova (see Bryzgunova, 1963, 1980; also Yanko, 2008) as "Intonational Construction 3"; it is often used for contrastive themes, and this is what takes place in this particular instance: One basket is contrasted to other elements of the previously introduced set of referents.

Generally, the specific direction of pitch is selected in accordance with the hierarchy "illocutionary function > EDU role in an illocutionary chain > role of an EDU-internal constituent" (Kibrik & Podlesskaya, 2009, pp. 95–96; cf. Kodzasov, 2009, pp. 103–104; Yanko, 2017). However, there is a considerable variation in pitch figures in particular accents. In Figure 4 the final EDU, a statement, predictably bears a falling pitch in the primary accent. In our transcription system, the statement-final role of this EDU is indicated with the period punctuation mark at the end. Other EDUs in Figure 4 are non-illocution-final, and that is indicated

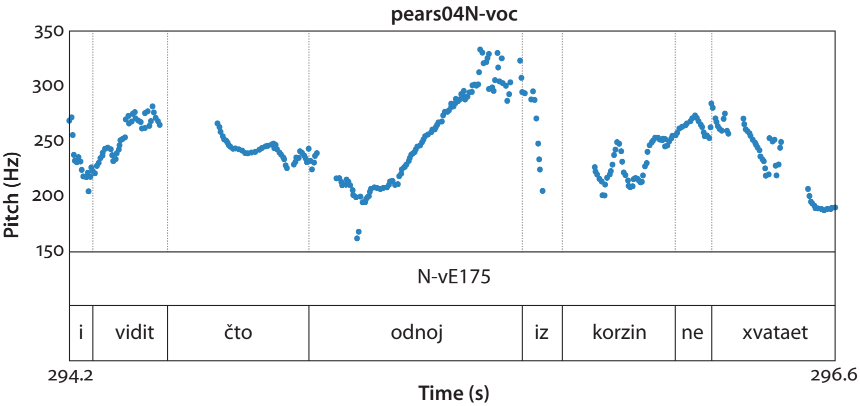


Figure 5. A rising and a falling accent in EDU N-vE175, example in Figure 4

with the commas. Specific prosodic manifestations of their primary accents (\↑ in all three instances) can be treated as instantiations of Bryzgunova's "Intonational Construction 4" (Bryzgunova, 1980); according to Yanko (2008, pp. 200–225), this pattern can be used in non-final elements of an unhurried narrative chain. The intonation contour, as it appears in N-vE172, can be seen in Figure 6. Note that this prosodic figure also applies to the increment noun phrase in N-vE174, which attributes this fragment the status of a legitimate member of the narrative chain. This prosodic figure is quite frequent in narrative discourse, although it differs from the most canonical comma intonation in Russian (see Section 6.1 below).

Spoken discourse contains analogs of what we know as written sentences. Such a *spoken sentence* is a sequence of EDUs implementing a particular illocutionary function. (As in common orthography, we capitalize the first letter of a spoken sentence.) The prosodic makeup of EDUs' primary accents differs crucially depending on whether the given EDU is the final or a non-final one in an illocutionary sequence, or a spoken sentence.[6]

Illocution-final EDUs are arranged in accordance with the phase of the illocutionary exchange level, while illocution-non-final EDUs follow the principles of illocution-internal phase. These two kinds of EDUs are considered in Sections 5 and 6, respectively.

[6] Our grouping of EDUs 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).

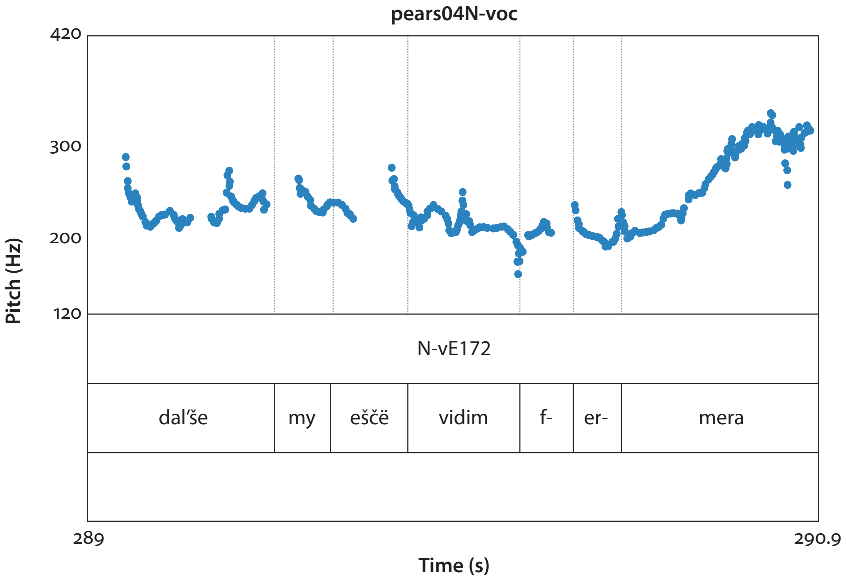


Figure 6. Intonational figure \↑ in EDU N-vE172, example in Figure 4

5. Illocution-final EDUs

5.1 Statement and question

Recall EDU N-vE175 in Figure 4 (cf. Figure 5). The primary accent of this illocution-final EDU bears a falling pitch accent targeting the very bottom of the speaker's f0 range. This is a typical instance of the "period intonation", encoding the completion of an illocutionary act of *statement*. As was pointed out above, this EDU also contains a thematic accent, adapting its direction of pitch to the final primary accent in a mirror-image way: As the anticipated rhematic accent is falling, the preceding thematic accent is rising.

The prosodic integrity of a statement EDU, supported by the mirror-image adaptation of a secondary accent to the primary accent, can be observed even in the instances of co-construction (e.g., see Lerner, 1991; Pekarek Doehler, 2011), when an EDU is produced via a joint effort of two interlocutors. See Figure 7, in which the Reteller and the Narrator jointly produce a simple clause with the illocutionary function of statement. The Reteller starts the EDU, uttering the thematic portion with a (secondary) rising accent. Then the Narrator snaps up, completing the EDU and the illocution and providing the rhematic portion with a primary falling accent.

TimeS	TimeE	Narrator	Reteller
543.62			R-vE027 /Šljapy tol’ko-o % hats only
544.39		N-vE264 % u \vzroslyx. at adults	
	544.66		
	545.06		

‘Hats, only adults have [them]’

Figure 7. Co-construction (source: pears22: R-vE027–N-v264)

(Instances of co-construction are marked in our transcription with the % symbols, closing and opening the two parts of a co-constructed sequence.)

Conversely to the statement illocution, the typical prosodic shape of a yes/no *question* is a rising primary accent; see Figure 8 and Figure 9. We use the ordinary question mark to transcribe this illocutionary function.

517.17	R-vE069	I	u	/mal’čika	tože	tak	*povjazano*?	Creaky voice during the middle portion of the accented vowel.
		and	at	boy	also	so	tied	

‘And is the boy’s [bandana] tied in the same way?’

Figure 8. A primary rising accent (source: pears04: R-vE069)

This example is somewhat unusual in that the primary accent is followed by nine unaccented syllables. This unaccented EDU-final sequence is a postponed theme; its final portion is characterized by high tempo of pronunciation (marked with italics in the transcript), which compensates for the unusual location of the primary accent.

In interactive discourse, turn-final statements sometimes contain a post-accent f0 rise (Kodzasov, 2009, pp. 109–110). One of the causes leading to this atypical behavior is that such statements, similarly to questions, may require a contribution from the interlocutor. An example of this kind can be seen in EDU N-vE223 in Appendix B. A somewhat similar phenomenon is discussed below in Section 5.3 under the label of semi-statement.

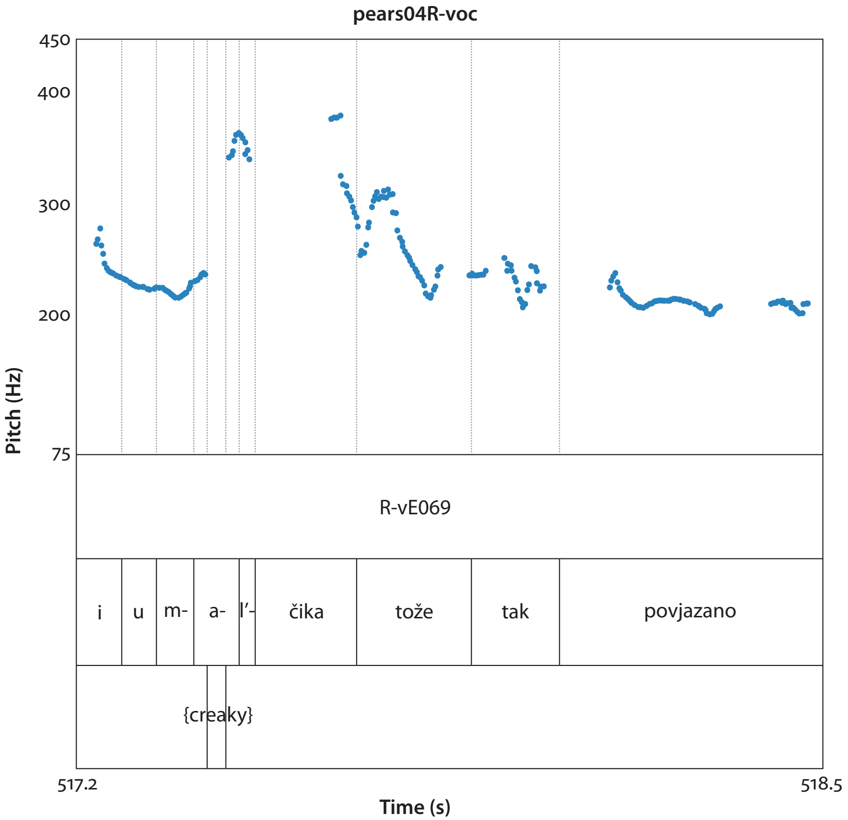


Figure 9. Rising primary accent marking a yes/no question in EDU R-vE169, example in Figure 8

5.2 Directive

We use the symbol ¡ to mark EDUs carrying the illocutionary function of *directive*. In the example in Figure 10, the Commentator reminds the Reteller about the existence of an additional character and suggests that she should not miss him in her ultimate retelling of the film.

666.23	C-vE109	Tam	eščë	byl	mužik	s	↑\kozoj.	
		there	also	was	man	with	goat	
667.81	C-vE110	Ty	ne	\zabud'	napisat'¡			In a low voice.
		you	not	forget	to.write			

'There was also the man with the goat. Don't forget to write about that.'

Figure 10. Directive (source: pears04: C-v109–C-vE110)

In Figure 10, the directive primary accent is conveyed with the falling pitch, although a rise is also possible (Kodzasov, 2009, pp. 106–107). Directives are infrequent in our data. From a structural point of view, they are in many ways similar to statements and questions. In particular, directive EDUs may contain secondary accents that adapt to the primary accents in their direction of pitch.

5.3 Semi-statement

Structurally similar to statements and questions is also a special illocution type, quite common in our data (particularly in the talk of Retellers), that we tentatively dub *semi-statements*. Formally, semi-statements resemble regular statements, as they are usually pronounced with a falling pitch accent. Functionally, however, a semi-statement is a request to confirm a guess; using this illocutionary type, a speaker looks forward to an immediate confirmation (or refutation) of his/her guess. Consider Figure 11, in which the Reteller is willing to ascertain the details regarding the goat in the film – the same one as mentioned in Figure 10. The primary accent in EDU R-vE132 is located on the rheme (*na-a* (0.13) \povodke 'on a lead') and bears a falling pitch (as opposed to a rise that would be expected in a regular yes/no question). However, functionally this utterance is much closer to a question. The Reteller does not have her own knowledge on the matter in question and unequivocally expects a reaction to her guess on the part of her interlocutors. Such a reaction, indeed, is immediately received from the Commentator, who cannot even wait until the end of the Reteller's semi-statement. We use the symbol ¿ to mark this illocutionary type.

Note that EDU R-vE132 also contains a secondary accent on the thematic constituent *ona* 'she'. In accordance with standard principles, it bears a rising pitch. (More precisely, a rising-level variety of it.)

TimeS	TimeE	Narrator		Reteller	
695.51				R-vE132	(ə 0.14) (ɯ 0.28) I-i /–ona-a n= || and she o= na-a (0.13) \povodke¿ on lead
698.09		C-vE129	(ɯ 0.28) \Da-a. yes		
	698.10				
	698.67				

R: 'And it [the she-goat] is on a lead [, right?]'
C: 'Yes'

Figure 11. Semi-statement (source: pears04 R-vE132 – C-vE129)

The functional similarity of questions and semi-statements is also manifested in a broader multichannel context. When requesting information, speakers may support their vocal actions with certain manual gestures and systematically rely on gaze direction to select an anticipated respondent; such non-vocal techniques are used both in questions and in semi-statements (see Korotaev, 2018a).

5.4 Vocative

In contrast to the above discussed illocutions, a *vocative* (alternative terms: address and allocution) does not presuppose a distinct communicative structure. When a vocative constitutes a separate EDU, we use the symbol @ to mark it. The prosodic makeup of vocatives is rather diverse and appears to follow principles different from those operating in the case of other illocutions. The example in Figure 12 is taken from the Night Dream Stories corpus. In this case the vocative is conveyed with a rise-fall accent (see Yanko, 2008, pp. 98–107, on the prosodic features of vocatives in spoken Russian).

25.23	E012	“/\Babuška@			
		grandma			
25.52	E013	A	–čto	èto	takoe?
		and	what	this	such
‘Grandma! What is that?’					

Figure 12. Vocative (source: Night Dream Stories, NDS027: E012–E013)

5.5 Exclamation

Exclamation is a special expressive and/or emphatic meaning. It is not an illocution as such, but rather an additional meaning, modifying the meanings of illocutions. When exclamation modifies a statement, we put a single symbol, the exclamation mark, at the end of the illocution-final EDU. This can be seen in EDUs N-vE228, N-vE229 and N-vE230 in Appendix B. Specifically, in N-vE230, the accent on *lestnice* ‘ladder’ bears a complex rising-falling pitch, which is typical of emphasis in Russian (see Yanko, 2008, pp. 83–97).

If an illocution other than statement contains the exclamatory meaning, double punctuation marks (e.g., ?! or ¡!) are used: The illocution mark is followed by the supplementary exclamation mark. Consider Figure 13 (its broader context can be seen in Appendix B). EDU R-vE025 contains the Reteller’s question, emphasized with an expressive meaning such as ‘I wonder’ or ‘hopefully’.

418.70	R-vE025	On	ne	/ly̱syj?!	High rising pitch.
		he	not	bald	
‘Is he bald?!’					

Figure 13. Exclamation (source: pears04: R-vE025)

To summarize this section, in our transcription system, illocution-final punctuation marks indicate discourse semantics rather than pure prosody. Prosody is annotated independently. In many instances there are standard ways to prosodically convey discourse functions, but this is not a one-to-one relationship.

6. Illocution-non-final EDUs

An EDU is considered illocution-final if it does not “project” (a term from Conversation Analysis; see Auer, 2005) a continuation, or, in a slightly different wording, it does not contain either verbal or prosodic “projectors”. Conversely, those EDUs that are interpreted as illocution-non-final contain verbal or prosodic signals implying the discourse meaning “to be continued”. In this section we review the main types of discourse *incompleteness*.

Discourse incompleteness may be more or less semantically loaded. There is a default type: incompleteness as such. This type of incompleteness is marked with a comma at the end of an EDU. Default incompleteness may be conveyed by several kinds of prosodic figures. Three kinds of such “comma intonations” are considered in Sections 6.1 to 6.3. Furthermore, there are other kinds of incompleteness, more special from a semantic point of view. They are described in Sections 6.4 to 6.6.

In our data, the majority of illocutions are statements, so the discussion below is confined to non-final statement EDUs.

6.1 Default incompleteness: Rising pitch accent

The most common type of comma intonation, used in the case of default incompleteness, is the rising primary accent. In the Night Dream Stories corpus, this kind of prosody accounts for 2/3 of all instances of comma-closed EDUs. The direction of pitch in these cases is guided by the principle of mirror-image adaptation introduced in Section 4 above: Since the anticipated statement-final EDU bears a falling pitch accent, non-final EDUs are equipped with rising pitch accents (cf. “the principle of melodic slope contrast” discussed by Debaisieux & Martin, this volume, Part I). In Figure 14 the non-final EDU R-vE340 is realized with a rising pitch accent (see Figure 15), mirroring the final fall in R-vE341. Other illustrations include, for example, the rising pitch accent in the first two EDUs in Figure 2 and EDU C-vE025 in Appendix B.

1215.71	R-vE340	a	ona	/upiraetsja,
		but	she	balks
1216.69	R-vE341	i-i	gromko	\bleet.
		and	loudly	bleats

‘But it [the she-goat] balks and bleats loudly’

Figure 14. Default incompleteness and rising pitch accent (source: pears04: R-vE340–R-vE341)

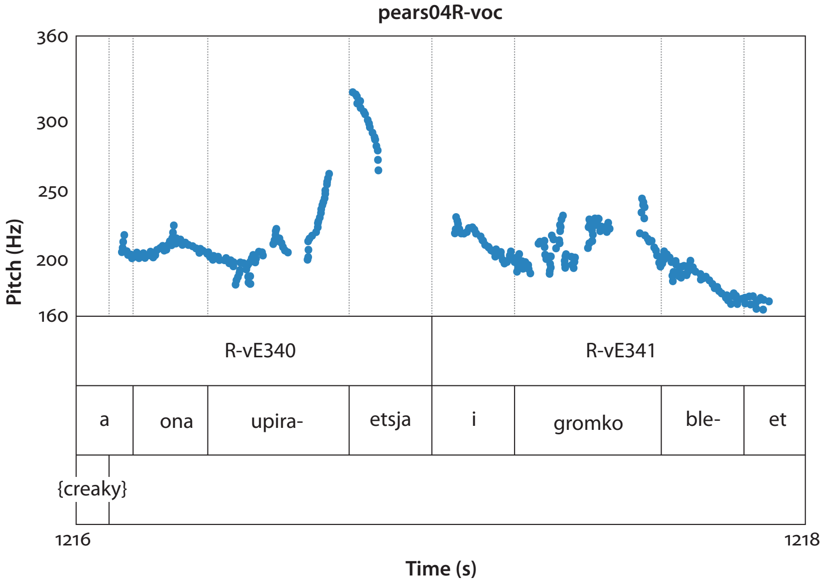


Figure 15. Rising pitch accent in EDU R-vE340, example in Figure 14

6.2 Default incompleteness: Falling pitch accent plus a subsequent rise

Quite common is also another prosodic figure we have already observed in the first three EDUs in Figure 4. In this case the EDU’s prosody is implemented with a fall-rise pattern, so that the primary accent bears a falling pitch, but the f0 curve rises further on; pitch changes direction either within the accented syllable (this is the only option if this is the final syllable of the EDU) or on subsequent syllables (Yanko, 2008, pp. 200–225). Figure 6 in Section 4 demonstrates how the f0 curve rises on post-accent syllables in EDU N-vE172.

Now consider the two EDUs in Figure 16 with the functionally identical prosodic patterns. In EDU N-vE202, the rise inevitably takes place within the accented syllable, as this is an EDU-final syllable. In EDU N-vE201, the accented syllable is also realized with the \/ figure, even though there are subsequent syllables; the rise continues steadily on these subsequent syllables. See Figure 17 for a visualization of this subtle distinction.

346.06	N-vE201	On \/↑polnen'kij, he chubby
346.75	N-vE202	u nego \/usy-y, at him moustache
'He is chubby, he's got a moustache…'		

Figure 16. Default incompleteness and fall-rise pitch accent (source: pears16: N-vE201–N-vE202)

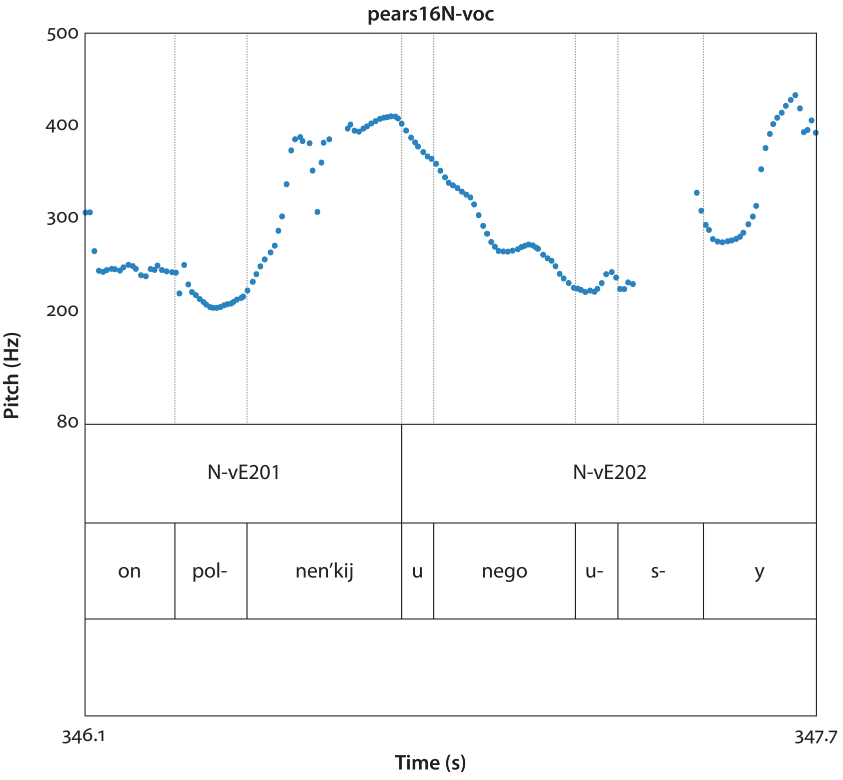


Figure 17. Fall-rise pitch accent in EDUs N-vE201 and N-vE202, example in Figure 16

6.3 Default incompleteness: Falling pitch accent

Apart from the fall-rise case considered in Section 6.2, there are instances in which a statement-non-final EDU is equipped with a simple falling pitch accent. This kind of the “falling comma” intonation differs from the period intonation in the target level of the speaker’s voice’s f0. Whereas in the period intonation the target level is about the absolute minimum of the given speaker’s f0 range, in the falling comma intonation the target level is about two or more semitones higher than that.

Two functional reasons may lead to the *non-final falling* phenomenon. First, the principle of mirror-image pitch adaptation may lead to the following interdependency: If a comma-closed EDU$_{n}$ bears a rising pitch in the primary accent (see Section 6.1 above), EDU$_{n\text{-}1}$ may bear the anticipatory falling pitch. An example can be seen in EDU C-vE024 in Appendix B.

Second, speakers sometimes use the gradual downstep strategy: When an upcoming period intonation is envisioned in EDU$_{n}$, a less low fall may be implemented in EDU$_{n\text{-}1}$ (and sometimes also in more than one EDU: EDU$_{n\text{-}2}$, EDU$_{n\text{-}3}$, etc.). In Figure 18, in the illocution-final EDU (N-vE061) pitch falls into the level of 170 Hz, while in the previous EDU (N-vE060) the f0 curve targets the level of 190 Hz, which is two semitones above; see Figure 19.

99.32	N-vE059	na-a	(0.19)	krasnom	/velosipe̲de,		
		on		red	bicycle		
100.96	N-vE060	kotoryj	emu	javno-o	(0.17)	\veli̲k,	Fall 190 Hz.
		which	him	obviously		too.big	
102.85	pN-027	(0.07)					
102.92	N-vE061	/očen’	emu	\bol’šo̲j.			Fall 170 Hz.
		very	him	big			
‘[the boy rides] a red bicycle, which is obviously too big for him, very big for him’							

Figure 18. Default incompleteness and non-final falling pitch accent (source: pears23: N-vE059–N-vE061)

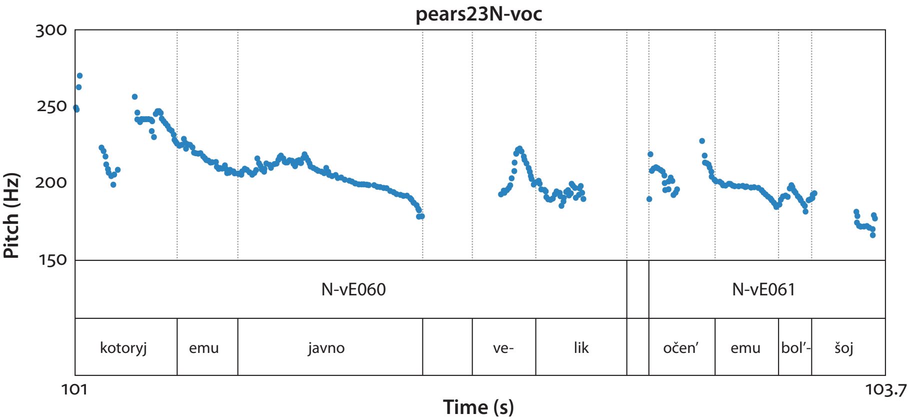


Figure 19. Downstep strategy in EDUs N-vE060 and N-vE061, example in Figure 18

6.4 Default incompleteness combined with a local illocutionary meaning

In all of the examples discussed in Sections 6.1 to 6.3, illocution-non-final EDUs have the same illocutionary semantics as the whole illocutionary sequence (or a "spoken sentence") they belong to, namely, they are statements. In such instances non-final EDUs are closed by a comma. However, there are instances in which non-final EDUs, prosodically belonging to a statement sentence, bear a different local illocutionary meaning. Such EDUs are closed by a double punctuation mark: an illocutionary symbol plus a comma. For example, in Figure 20 the speaker, who is also the main character of the story, gives a command to other characters in the non-final EDU E026 and then completes the statement-type sentence. The phenomenon of illocutionary heterogeneity in spoken Russian was previously noticed in Zemskaja, Kitajgorodskaja, & Širjaev (1981).

47.70	E025	“\La-adno,				
		OK				
48.00	E026	vy	ex= ||	exajte	bez	/menja¡,
		you.guys	g=	go.IMPER	without	me
49.10	E027	ja	potom	\s-sama	priedu.	
		I	later	myself	will.come	
‘OK, you guys go [take the elevator] without me, I will come later myself’						

Figure 20. Incompleteness and local illocutionary semantics (source: Night Dream Stories, NDS050: E025–E027)

6.5 Elucidation

In Section 6.3 we discussed non-illocution-final EDUs with the falling pitch in the primary accent, but bearing the meaning of default incompleteness. Semantically more special is the context of elucidation, also associated with the falling pitch in the primary accent. The cover term *elucidation* embraces two particular semantic contexts: cataphoric introduction and quotation. We consider them in turn. In both cases EDUs in question are closed with the colon.

In the case of cataphoric introduction, the first, colon-closed, EDU contains a cataphoric element elucidated in the subsequent EDU(s). A typical example can be seen in Figure 21.

953.14	R-vE189	Scena \takaja: scene such
954.21	R-vE190	est' /xolm, there.is hill
955.21	pR-120	(0.45)
955.66	R-vE191	gde-to po centru vdali /gory, somewhere in center far mountains
957.85	pR-121	(0.37)
958.22	R-vE192	na perednem plane (ə 0.33) odno || odno /grúševoe \derevo. on fore ground one one pear tree
'The scene is as follows: there is a hill; somewhere in the middle, at a distance, there are mountains; in the foreground, there is one… one pear tree'		

Figure 21. Elucidation (source: pears16: R-vE189–R-vE192)

A cataphoric element may be formally missing, while semantically the cataphoric relation between the introductory, colon-marked, EDU and the subsequent EDU(s) is still in place. For example, see Figure 22.

447.45	C-vE046	/Korzin tam bylo \tri: baskets there were three
448.49	C-vE047	dve /polnye, two full
449.32	C-vE048	odna \pustajaˀ. one empty
'There were three baskets there: two full ones, [and] an empty one'		

Figure 22. Missing cataphoric element (source: pears04: C-vE046–C-vE048)

A further example is found in Appendix B: The colon-marked EDU C-vE020, saying that the farmer's clothing was atypical, is elucidated in the sequence C-vE021 to C-E026, providing details concerning the farmer's clothes.

The second context of elucidation is direct quotation. The first, colon-closed, EDU introduces the subsequent quoted speech or thought. A direct quote can be distinguished from an indirect quote with the help of a number of criteria, including the character of deictic elements and the presence of main clause phenomena (Aelbrecht, Haegeman, & Nye, 2012). We transcribe direct quotes using regular "quote marks".

For example, in Figure 23 EDUs N-vE491 and N-vE492 constitute a direct quote; this is confirmed by the first person deixis and by the presence of an interjection. The latter is an unequivocal representative of main clause phenomena: There are no ways to convey interjections in the indirect speech mode. EDU N-vE490 is the speaker's introductory device, in this case represented not by a verb of speech, but by the adjectival demonstrative *takoj* 'such', analogous to English *kind of* or *like*, sometimes treated as "new quotatives" (Buchstaller, 2014; Buchstaller & van Alphen, 2012). The demonstrative is produced as a prosodically autonomous EDU, bearing its own primary accent.

910.55	N-vE489	Tot /bežit, that.one runs
910.99	N-vE490	\takoj: such
911.28	N-vE491	"/\ Xè-èj! hey
911.55	N-vE492	U menja /→ gruši", at me pears
912.08	N-vN043	(ɥ 0.34)
912.42	N-vE493	/\ vot. well
912.89	N-vE494	Vsem /daët, to.all gives
913.69	N-vF035	(ˀ 0.29) (ə 0.60) (ˀ 0.10)
914.68	pN-271	(0.21)
914.90	N-vE495	i oni /idut \dal'še. and they go further

'That one runs [and shouts], like: "Hey! I've got pears!", [and] well, [he] gives [pears] to everyone and they go further'

Figure 23. Direct quote (source: pears16: N-vE489–N-vE49)

The common denominator of the cataphoric and quotative construction is the presence of a projector, foretelling the appearance of a subsequent discourse fragment that specifies the introductory, colon-marked, EDU. The falling pitch in the primary accent of introductory EDUs creates a unique mismatch between meaning and prosody: The prosodic implementation by itself does not project a continuation, while the cataphoric semantics or a quote-introducing device, on the contrary, implies a subsequent specification.

6.6 Inexhaustiveness

In Russian spoken discourse, there is a frequent prosodic figure, in which a primary accent contains a moderate pitch rise, followed by a level f0 interval or a minor fall on the subsequent syllables; often a level period is observed already within the accented syllable. This prosodic figure is known as "Intonational Construction 6" in Bryzgunova's system (Bryzgunova, 1980). The function of this figure, particularly when the accented syllable is lengthened, is to demonstrate mental activity in situations where information is partly missing, such as trying to recollect something or pondering possible alternatives. In addition, it is used in describing an open list of events or objects (Yanko, 2008, pp. 109–113, 166–167). The term *inexhaustiveness* is a tentative umbrella label for this range of meanings.

When hearing an EDU with such a prosodic arrangement, a listener may suppose that a continuation could follow, possibly with the same prosodic figure. However, this is usually less than clear from the prosody of the given EDU. If a continuation actually follows, we use the punctuation mark,,, at the end of the EDU, and if the given illocutionary sequence is completed, the symbol … is used.

In the example in Figure 24, the inexhaustiveness strategy is used three times: in EDUs R-vE294, R-vE296 and R-v299. Every time the speaker makes the decision to go on with the current illocutionary sequence, so the three commas symbol is used in all of these EDUs. It is quite common that the inexhaustiveness strategy is kept throughout a group of EDUs; by using this prosodic strategy, the speaker enters a certain mental mode that does not have to be abandoned right away. Figure 25 illustrates the f0 curve in EDU R-vE296. The accented syllable demonstrates a moderate rise, followed by an interval of slow falling.

In Appendix B, the ,,, symbol is found in EDUs C-vE021, C-vE022, and C-vE029. The same prosodic pattern occurs in EDU C-vE030, but, in contrast to the previous instances, it is found in an illocution-final EDU, so the transcript shows the three dots symbol. The Commentator's EDU sequence is conveyed as being potentially extendable, but still gives the interlocutors an opportunity for turn-taking, which both of them immediately embrace: The Narrator expresses her agreement in EDU N-vE229, while the Reteller asks a confirmation question and then goes on with a

1142.17	R-vE293	\Vot, well
1142.61	R-vE294	(əɯ 0.58) u nego /lestnica →pristavlena,,, at him ladder leaned
1144.55	R-vE295	(k –derevu,) to tree
1145.25	R-vN021	(ɥ 0.58)
1145.84	R-vE296	(ə 0.27) i \on /–zalezaet,,, and he climbs.up
1147.36	pR-234	(0.44)
1147.79	R-vE297	(ɯ 0.17) (–vot,) well
1148.17	R-vE298	on zalezaet na /lestnicu, he climbs.up on ladder
1149.12	R-vE299	ona /↓poskripyvaet,,, it squeaks
1150.10	R-vE300	a on sobiraet \gruši. and he picks pears

‘Well, he has a ladder leaning at the tree, and he climbs up, well, he climbs up the ladder, it squeaks, and he is picking pears’

Figure 24. Inexhaustiveness strategy (source: pears04: R-vE293–R-vE300)

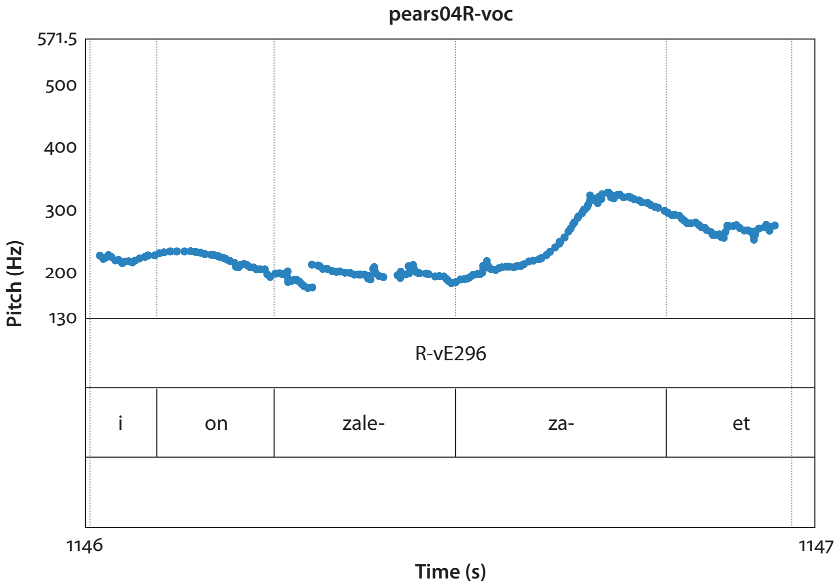


Figure 25. Inexhaustiveness intonational pattern in EDU R-vE296, example in Figure 24

close duplicate of the Commentator’s EDU, imitating even his prosody (R-vE024, also closed by three dots).

As Yanko (2008) points out, the prosodic strategy of inexhaustiveness is sometimes backed with non-vocal actions. In particular, some speakers reinforce the idea of an ongoing mental activity by using particular head movements.

The notion of inexhaustiveness may combine with illocutions other than statement. In this case the main illocutionary semantics and the inexhaustiveness interpretation are identified separately, on the basis of the EDU’s verbal content, prosody, and the context. For example, consider EDU R-vE021 in Appendix B. The primary accent on the verb (*/ispo̲l’zuet*) is typical for a yes/no question, while the special inexhaustiveness prosody is expressed on the subsequent direct object (*/→le-estnicu*); also note the discourse marker *tam* ‘or something like that’, additionally pointing to insufficient confidence. Accordingly, the symbol combination ?… is used in the transcript. See also EDU R-vE022 in Appendix B, where inexhaustiveness is combined with the inverse question mark, indicating a semi-statement.

To summarize this section, a punctuation mark at the end of an illocution-non-final EDU codes the appropriate discourse function, while prosody is transcribed independently. In most instances, specific prosodic figures used to encode a particular discourse function are highly limited and partly, though not fully, predictable.

7. Disfluencies

While producing discourse, speakers may experience various kinds of difficulties. Consider the first EDU in Figure 4 (EDU N-vE172). At that point, the Narrator apparently has difficulties in recounting the beginning of a film episode. These difficulties surface as certain vocal phenomena: filled *hesitation* pauses (ˀ 0.32) and (ə 0.69), as well as the lengthening of the EDU-initial conjunction (*i-i*). The following EDU, N-vE173, also begins with a hesitation pause (ɐ 0.11). Such vocal elements, sometimes attributed a lexical status (see Clark & Fox Tree, 2002), are used when speakers are not yet ready to contribute a verbalization of their thought they would find satisfactory, and require some extra time to ponder on how to progress, at the same time signaling the interlocutor that they are willing to go on.

In contrast to absolute (silent) pauses, the attribution of filled pauses to a particular speaker is quite obvious, hence, in a scores transcript, filled pauses are annotated within individual speakers’ columns. We distinguish between four kinds of sounds that can fill hesitation pauses: vowels (ə) and (ɐ), the glottal approximant (ˀ), and the nasal sonorant (ɯ). These symbols are written in parentheses and are followed by a number indicating the sound’s length in seconds. There are also mixed sounds, as for the example in Figure 24, where EDU R-vE294 starts with the filled pause (əɯ). Hesitation pauses do not have to be filled, they can also be silent; for example, see such a silent pause in the middle of EDU N-vE060 in Figure 18. Silent boundary pauses can also include a hesitation component, but that is generally undetectable from the vocal signal as such.

Let us go back to the example in Figure 4. As has been already mentioned, EDU N-vE172, along with filled pauses, also contains the lengthening of a lexical vowel. Generally, a lengthening of an EDU-initial conjunction/complementizer is often combined with a subsequent filled pause or another lengthening. Hesitation-related lengthening most often affects word-initial and word-final phonemes, both vowels and consonants; see examples in N-vE175 (Figure 4), N-vE192 (Figure 16), N-vE057 (Figure 18), or E027 (Figure 20). Hesitation-related lengthening should be distinguished from other possible cases of sound lengthening, as for example, emphatic or connected with the phenomenon of inexhaustiveness discussed in Section 6.6 above.

From the perspective of production, hesitation can be seen as an early detected speech disfluency: A speaker realizes a certain problem and applies effort towards mending it before a problematic element has been verbalized. Accordingly, from a structural point of view, hesitation is a relatively mild kind of disfluency. More severe disfluencies take place when a speaker has already begun or even completed an unsatisfactory verbal element and has to drop it and, possibly, replace it with something different. It is not infrequent that a word is begun, but then is interrupted and remains unfinished. In the case of such truncation we use the = symbol; see EDU E026 in Figure 20, where the speaker first only pronounces the beginning of the word (*ex*=) and says the word in full afterwards (*exajte*). (The same = symbol is used when a word is interrupted and then resumed, cf. the disrupted word, pronounced in three pieces, in EDU C-vE071 in Figure 27 below: \obloko= *ti̱v*= *šis'*.)

The relatively late and severe disfluencies, associated with rejecting something already said, are conventionally called *false starts* or *repairs*. False starts fall into two major categories depending on the scope of repair: The criterion is whether the speaker manages to complete the current EDU in spite of his/her difficulties or the speaker abandons the EDU under construction. This distinction is cross-cut by a second parameter: whether the false start happens because of the speaker's own internal cognitive processes or because of the pressure coming from external causes. Internally-induced false starts are marked with the following symbols: || in the case of an EDU that was eventually repaired, and == if an EDU was abandoned. Similarly, externally-induced false starts are marked by the symbols ≀≀ and ≈≈.

The disfluency in the above discussed EDU E026, in Figure 20, is local, it is immediately repaired. The disfluency found in Figure 26 is also due to a local problem: The speaker needs to replace the plural pronoun with the singular one. The speaker discovers this problem after he has already begun producing the verb form, so he has to abandon it and go one step back, now choosing the correct pronoun and recycling the beginning of the verb form.

1045.49	C-vE226	/Poètomu	v	tot	moment	kogda	oni	/pada= ||	(ˀ 0.10)	kogda
		that.is.why	at	that	moment	when	they	fa=		when
		on	/padaet,							
		he	falls							
‘That’s why at that [very] moment when they fa- …. when he falls…’										

Figure 26. Repair (source: pears04: C-vE226)

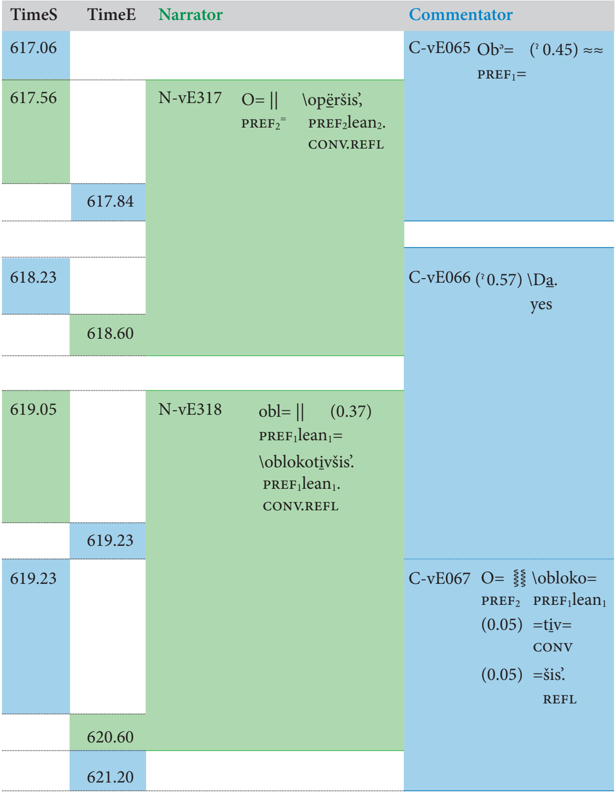
A more severe difficulty that the speaker cannot resolve within the bounds of an EDU is found in EDUs R-vE016–R-vE017 in Appendix B. When formulating her question, the Reteller first tries to use the wh-word *čto* ‘what’, but after two truncated attempts to say this word abandons the original plan and ends up using a yes/no, rather than a wh-question. In accordance with the above-mentioned transcription principles, the first false start is transcribed with the help of the EDU-internal || symbol, while the second one with the help of the EDU-final == symbol.

In the example in Figure 27 both interlocutors jointly engage in overcoming the difficulties associated with pronouncing the converb *oblokotivšis’* ‘having leaned upon’.[7] The Commentator makes the first attempt in EDU C-vE069, but only manages to pronounce the prefix *ob*= (the adventitious sound ᵊ appears at the moment of interruption). At this moment the Narrator jumps in; in EDU N-vE338 she first says the prefix *o*= and then offers the synonymous converb *opëršis’* ‘having rested himself upon’. In EDU C-vE070 the Commentator expresses her agreement with this choice. The Narrator, however, seems to be unsatisfied herself with her version and the following EDU N-vE339 goes back to the form *oblokotivšis’* (again with a false start, on the first attempt only saying the initial part of the word). Now, in EDU C-vE071 the Commentator first proceeds with the beginning of *opëršis’*, then continues with the form last contributed by the Narrator, even though she has hard time producing this form: She interrupts the word twice with brief pauses.

As is clear from the above discussed examples, at the interruption points of the repairs we often see hesitation markers; cf. filled creaky hesitation pauses in EDUs C-vE226, considered above in Section 5.3, there is both a repair (*n*= is replaced by *na-a povodke*) and a whole gamut of hesitation markers: two initial filled pauses of different kinds and three instances of vowel lengthening.

For further details of our approach to speech disfluencies see Podlesskaya (2015). In Appendix B various kinds of disfluencies are found in EDUs R-vF002, R-vE016, R-vE017, C-vE019, R-vE019, R-vE021, C-vE020, C-vE021, C-vE025, N-vF009, C-vE029, C-vE031, N-vE232, and C-vE033.

[7] To make this example more transparent for non-Russian speakers, we provide a more detailed morphological glossing than in other cases: Added abbreviations are PREF for a derivational prefix, CONV for converb, REFL for reflexive.



TimeS	TimeE	Narrator	Commentator
617.06			C-vE065 Obᵊ= (ˀ 0.45) ≈≈ PREF$_1$=
617.56		N-vE317 O= || \opëršis’, PREF$_2$= PREF$_2$lean$_2$. CONV.REFL	
	617.84		
618.23			C-vE066 (ˀ 0.57) \Da. yes
	618.60		
619.05		N-vE318 obl= || (0.37) PREF$_1$lean$_1$= \oblokotivšis’. PREF$_1$lean$_1$. CONV.REFL	
	619.23		
619.23			C-vE067 O= ≀≀ \obloko= PREF$_2$ PREF$_1$lean$_1$ (0.05) =tiv= CONV (0.05) =šis’. REFL
	620.60		
	621.20		

C: ‘Having …’
N: ‘H- having rested himself upon’
C: ‘Yes’
N: ‘having… leaned upon’
C: ‘H- having lea-n-ed upon’

Figure 27. Collaborative repairs (source: pears16: C-vE069–C-vE071)

8. Other phenomena

A fairly common phenomenon of local discourse structure is the *inset*. Suppose there is a sequence EDU$_1$ + EDU$_2$, such that both of these EDUs semantically belong to the mainline of the discourse. There may be an inset between EDU$_1$ + EDU$_2$, not belonging to the mainline and providing supplementary, optional, or background information. We transcribe insets with the help of parentheses, see for example EDU C-vE023 in Appendix B. In this case the inset consists of one EDU and is a part of a larger sentence. None of these features is obligatory: Insets can embrace series of EDUs and can represent whole sentences.

An inset may wedge in not between two EDUs but between two parts of one and the same EDU. This is the structure we call *split*. Split is transcribed with the em-dashes at the end of the first part of the interrupted EDU and at the beginning of its second part. In Figure 28, the first part of the EDU contains the first conjunct of a conjoined NP, the inset is a restrictive relative clause specifying this conjunct, and the second part of the EDU involves the second conjunct.

359.13	C-vE014	Kstati na vot ètom /mal'čike — by.the.way on here this boy
360.63	C-vE015	(kotoryj na \velike ezdilˀ,) which on bike rode
362.05	C-vE016	— ˀi na-a /djad'ke /fermere byli platki \↑odinakovye! and on man farmer were scarves same
'By the way, the scarves worn by this very boy (the one who rode the bicycle) and by the farmer guy were alike'		

Figure 28. Split (source: pears04: C-vE014–C-vE016)

In our analysis of spoken discourse, we address a number of further structural and prosodic phenomena, including emphasis, accelerated tempo, lowered f0 register, reduction, soft pronunciation, laugh, and so forth. There is no space to discuss these phenomena here, but see Appendix A for a complete list of transcription conventions. Discussion and examples can be found in Kibrik & Podlesskaya (2009) (in Russian) and in Fedorova & Kibrik (2020), as well as at the web sites <spokencorpora.ru> and <multidiscourse.ru>. One point should be noted here. Sometimes, EDUs may contain a significant internal seam, that is, a potential place for further segmentation. In this chapter, as well as in Appendix B, there are no instances of this phenomenon, but see Kibrik, Korotaev, & Podlesskaya (this volume, Part II, Section 3) for a discussion based on English evidence.

9. Conclusion: Vocal channels and their interaction with non-vocal channels

We have thus reviewed the most prominent phenomena, associated with the local structure and prosody of Russian spoken discourse. As was pointed out in Section 1 above, we now approach talk in a broader context of multichannel communication (see Adolphs & Carter, 2013; Kress, 2010; Mondada, 2016; Müller et al., 2013; inter alia, on the contemporary multimodal/multichannel agenda). In addition to the vocal modality, there is a multiplicity of channels and components of the kinetic modality, including eye gaze, manual gestures and other kinds of gestures (including face, head, and torso gestures). This perspective calls for a more encompassing theory of communicative behavior. A preliminary version of such theory is presented in Kibrik (2018), and here we briefly mention some of the crucial points.

In the tradition restricted to vocal discourse, there is a sharp distinction between production and comprehension, between the roles of speaker and listener. However, in multichannel communication the speaker simultaneously monitors the listener's kinetic behavior, such as gaze and gestures, and this affects how speech proceeds. That is, the addressant is also an addressee, and production and comprehension take place at the same time. Furthermore, the notion of turn becomes more complex and less discrete if kinetic behavior is taken into account. The notion of absolute pause also becomes tricky, as communicative behavior never stops. Even a static posture assumed by a listener is a kind of an informative signal sent to the speaker and is taken sometimes as an incentive to keep going with the talk.

Our notion of elementary discourse unit calls for specification, given that discourse is not just vocal. There are similar units in other channels of behavior. In particular, manual gestures are comparable to vocal EDUs in many ways. In fact, they can be considered manual EDUs. When communication is discussed in a multichannel perspective, a unit of vocal behavior should probably be recast as a *vocal elementary discourse unit* (vEDU).

As was discussed in Section 3.2, EDUs correlate with clauses. There is substantial literature on the relationship between gestures and clauses (e.g., McNeill, 1992). We have explored the temporal coordination between EDUs and manual gestures in our data (Fedorova et al. 2016; Korotaev, 2018b). These studies generally confirm the tendency of such temporal coordination.

Multimodal/multichannel studies often confine their perspectives to a relationship between the verbal and the kinetic (particularly gestural) structure. In our approach, prosody plays a very important role in the interaction of communication channels. In many ways prosody is a bridge between the verbal structure and the kinetic behavior. In particular, there is a significant similarity between the prosodic phenomenon of primary accent and the stroke phase in gestures. In this sense the material of this chapter may be useful not only for purely vocal studies, but also for a broader range of explorations in human communication.

Acknowledgements

The authors are grateful to the editors of this book for their comments on various drafts of this paper, as well as for inventing this collection and putting it together.

References

Adolphs, S., & Carter, R. (2013). *Spoken corpus linguistics: From monomodal to multimodal*. New York, NY: Routledge. https://doi.org/10.4324/9780203526149

Aelbrecht, L., Haegeman, L., & Nye, R. (Eds.). (2012). *Main clause phenomena: New horizons*. Amsterdam: John Benjamins. https://doi.org/10.1075/la.190

Auer, P. (2005). Projection in interaction and projection in grammar. *Text*, 25(1), 7–36. https://doi.org/10.1515/text.2005.25.1.7

Bryzgunova, E. A. (1963). *Praktičeskaja fonetika i intonacija russkogo jazyka* [Practical phonetics and intonation in Russian]. Moscow: MSU.

Bryzgunova, E. A. (1980). Intonacija [Intonation]. In N. Švedova, A. Bondarko, V. Ivanov, V. Lopatin, I. Uluhanov, N. Arutyunova (Eds.), *Russkaja grammatika* [Russian grammar] (Vol. 1, pp. 98–118). Moscow: Nauka.

Bossaglia, G., Mello, H., & Raso, T. (this volume). Illocution as a unit of reference for spontaneous speech: An account for insubordinated adverbial clauses in Brazilian Portuguese. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Buchstaller, I. (2014). *Quotatives: New trends and sociolinguistic implications*. Malden, MA: Wiley-Blackwell.

Buchstaller, I., & van Alphen, I. (Eds.). (2012). *Quotatives: Cross-linguistic and cross-disciplinary perspectives*. Amsterdam: John Benjamins. https://doi.org/10.1075/celcr.15

Chafe, W. (Ed.). (1980). *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.

Chafe, W. (1994). *Discourse, consciousness, and time*. Chicago, IL: University of Chicago Press.

Clark, H. H., & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, 84, 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Cresti, E. (this volume). The pragmatic analysis of speech and its illocutionary classification according to Language into Act Theory. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Croft, W. (1995). Intonation units and grammatical structure. *Linguistics*, 33, 839–852. https://doi.org/10.1515/ling.1995.33.5.839

Debaisieux, J.-M., & Martin, P. (this volume). Syntactic and prosodic segmentation in spoken French. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1992). Discourse transcription. *Santa Barbara Papers in Linguistics*, 4, 1–225.

Fedorova, O.V., & Kibrik, A.A. (Eds.) (2020). *The MCD handbook: A practical guide to annotating multichannel discourse*. Moscow: Institute of Linguistics RAS.

Fedorova, O. V., Kibrik, A. A., Korotaev, N. A., Litvinenko, A. O., & Nikolaeva, Ju. V. (2016). Vremennaja koordinacija meždu žestovymi i rečevymi edinicami v mul'timodal'noj kommunikacii [Temporal coordination between gestural and speech units in multimodal communication]. *Computational Linguistics and Intellectual Technologies: Papers From the Annual International Conference "Dialog"*, 15 (22), 159–170. Retrieved from <http://multidiscourse.ru/data/pub/fedorova%20et%20al%20dialog%202016.pdf>

Fernandez-Vest, M. M. J. (2016). Detachment linguistics and information grammar of oral languages. In M. M. J. Fernandez-Vest & R. D. Van Valin , Jr. (Eds.), *Information structuring of spoken language from a cross-linguistic perspective* (pp. 7–32). Berlin: De Gruyter Mouton.

Ford, M., & Holmes, V. M. (1978). Planning units and syntax in sentence production. *Cognition*, 6(1), 35–53. https://doi.org/10.1016/0010-0277(78)90008-2

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech*, 15, 103–113. https://doi.org/10.1177/002383097201500201

Izre'el, S. (this volume). The basic unit of spoken language and the interface between prosody, discourse and syntax: A view from spontaneous spoken Hebrew. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Iwasaki, S., & Tao, H. (1993). *A comparative study of the structure of the intonation unit in English, Japanese, and Mandarin Chinese*. Paper presented at the Annual Meeting of the Linguistics Society of America, Los Angeles, CA. Retrieved from <https://www.researchgate.net/publication/241563041_A_Comparative_Study_of_the_Structure_of_the_Intonation_Unit_in_English_Japanese_and_Mandarin_Chinese>

Jaworski, A. (Ed.). (1997). *Silence: Interdisciplinary perspectives*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110821918

Jefferson, G. (2004). Glossary of transcript symbols. In G. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 13–31). Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.125.02jef

Kibrik, A. A. (2011). Cognitive discourse analysis: Local discourse structure. In M. Grygiel & L. A. Janda (Eds.), *Slavic linguistics in a cognitive framework* (pp. 273–304). Frankfurt: Peter Lang.

Kibrik, A. A. (2015). *Pear stories, 40 years later*. Symposium conducted at the EuroAsianPacific Joint Conference on Cognitive Science, Torino, Italy, September 2015. Retrieved from <http://ceur-ws.org/Vol-1419/section0010.pdf>

Kibrik, A. A. (2018). Russkij mul'tikanal'nyj diskurs. Čast' II. Razrabotka korpusa i napravlenija issledovanij [Russian multichannel discourse. Part II. Corpus development and avenues of research]. *Psixologičeskij Žurnal*, 39(2), 79–89.

Kibrik, A. A., & Fedorova, O. V. (2018). An empirical study of multichannel communication: Russian pear chats and stories. *Psychology. Journal of the Higher School of Economics*, 15(2), 191–200.

Kibrik, A. A., Korotaev, N. A., & Podlesskaya, V. I. (this volume). The Moscow approach to local discourse structure: An application to English. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Kibrik, A. A., & Podlesskaya, V. I. (2006). Problema segmentacii ustnogo diskursa i kognitivnaja sistema govorjaščego [Segmentation of spoken discourse and the speaker's cognitive system]. In V. D. Solovyev (Ed.), *Kognitivnye issledovanija* [Cognitive studies], 1 (pp. 138–158). Moscow: Institut psixologii RAN.

Kibrik, A. A., & Podlesskaya, V. I. (Eds.). (2009). *Rasskazy o snovidenijax: korpusnoe issledovanie ustnogo russkogo diskursa* [Night dream stories: A corpus study of spoken Russian discourse]. Moscow: Jazyki slavjanskix kul'tur.

Kodzasov, S. V. (1996). Kombinatornaja model' frazovoj prosodii [A combinatory model of phrasal prosody]. In T. M. Nikolaeva (Ed.), *Prosodičeskij stroj russkoj reči* [Prosodic structure of the Russian speech] (pp. 85–123). Moscow: IRJA RAN.

Kodzasov, S. V. (2009). *Issledovanija v oblasti russkoj prosodii* [Studies in the field of Russian prosody]. Moscow: Jazyki slavjanskix kul'tur.

Korotaev, N. A. (2015). Kommunikativno-prosodičeskij podxod k vyjavleniju èlementarnyx diskursivnyx edinic v ustnom monologičeskom tekste [Elementary discourse units in spoken monologues: Evidence from communicative prosody]. *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialog"*, 14(21), 294–307.

Korotaev, N. A. (2018a). Vopos i poluutverždenie v structure mul'tikanal'nogo diskursa [Questions and semi-statements in multichannel discourse]. In A. K. Krylov & V. D. Solovyev (Eds.), *The eigth international conference on cognitive science. October 18–21, 2018, Svetlogorsk, Russia. Abstracts* (pp. 1311–1313). Moscow: Institut psixologii RAN.

Korotaev, N. A. (2018b). O vremennoj koordinacii žestikuljacionnyx i rečevyx edinic v nepodgotovlennoj ustnoj kommunikacii [On temporal coordination between gesture and speech units in spontaneous spoken discourse]. In S. O. Savčuk (Ed.), *"Slovo i žest"* [Speech and gesture] (pp. 10–12). Moscow: Institut russkogo jazyka.

Kress, G. (2010). *Multimodality: A social semiotic approach to communication*. London: Routledge Falmer.

Krivnova, O. F. (2007). Faktor rečevogo dyxanija v intonacionno-pauzal'nom členenii reči [The factor of speech breathing in the intonational-pausal segmentation of speech]. In V. A. Vinogradov (Ed.), *Lingvističeskaja polifonija. Sbornik v čest' jubileja prof. R. K. Potapovoj* [Linguistic polyphony. Festschrift to honor R. K. Potapova] (pp. 424–445). Moscow: Jazyki slavjanskix kul'tur.

Lerner, G. H. (1991). On the syntax of sentences-in-progress. *Language in Society*, 20, 441–458. https://doi.org/10.1017/S0047404500016572

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

Markus, O. V. (2009). Lokal'naja struktura diskursa v verxnekuskokvimskom atabaskskom jazyke [Local discourse structure in Upper Kuskokwim Athabaskan] (Unpublished doctoral dissertation). Lomonosov MSU, Moscow, Russia.

Maruyama, T., Den, Y., & Koiso, H. (this volume). Design and annotation of two-level utterance-units: From Japanese viewpoint. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Matsumoto, K. (2003). *Intonation units in Japanese conversation: Syntactic, informational and functional structures*. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.65

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics*, 20, 336–366. https://doi.org/10.1111/josl.1_12177

Müller, C., Cienki, A., Fricke, E., Ladewig, S. H., McNeill, D., & Teßendorf, S. (Eds.). (2013). *Body – Language – Communication: An international handbook on multimodality in human interaction*. Berlin: De Gruyter Mouton.

Pawley, A., & Syder, F. H. (2000). The one-clause-at-a-time hypothesis. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 163–199). Ann Arbor, MI: University of Michigan Press.

Pekarek Doehler, S. (2011). Emergent grammar for all practical purposes: The on-line formatting of left and right dislocations in French conversation. In P. Auer & S. Pfänder (Eds.), *Constructions: Emerging and emergent* (pp. 45–87). Berlin: De Gruyter. https://doi.org/10.1515/9783110229080.45

Podlesskaya, V. I. (2011a). Relative clauses in spoken Russian and elsewhere: A corpus approach. *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialog"*, 10(17), 529–537.

Podlesskaya, V. I. (2011b). K tipologii strukturnyx edinic v spontannoj ustnoj reči [Towards a typology of structural units in spontaneous spoken discourse]. In *Meždunarodnaja konferencija, posvjaščennaja 50-letiju Peterburgskoj tipologičeskoj školy* [Fifty years of St. Petersburg school of linguistic typology: Papers from the international conference] (pp. 146–150). St. Petersburg: Nestor-Istorija.

Podlesskaya, V. I. (2015). A corpus-based study of self-repairs in Russian spoken monologues. *Russian Linguistics*, 39(1), 63–79. https://doi.org/10.1007/s11185-014-9142-1

Sornicola, R. (2006). Interaction of syntactic and pragmatic factors on basic word order in the languages of Europe. In G. Bernini & M. L. Schwartz (Eds.), *Pragmatic organization of discourse in the languages of Europe* (pp. 357–544). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110892222.357

Švedova, N., Bondarko, A., Ivanov, V., Lopatin, V., Uluhanov, I., Arutyunova, N. (Eds.). (1980). Russkaja grammatika [*Russian grammar*] (Vol. 1). Moscow: Nauka.

Tannen, D., & Saville-Troike, M. (Eds.). (1985). *Perspectives on silence*. Norwood, NJ: Ablex.

Thompson, S. A., & Couper-Kuhlen, E. (2005). The clause as a locus of grammar and interaction. *Discourse Studies*, 7, 481–506. https://doi.org/10.1177/1461445605054403

Wouk, F. (2008). The syntax of intonation units in Sasak. *Studies in Language*, 32, 137–162. https://doi.org/10.1075/sl.32.1.06wou

Yanko, T. (2008). *Intonacionnye strategii russkoj reči v tipologičeskom aspekte* [Intonational strategies in spoken Russian from a comparative perspective]. Moscow: Jazyki slavjanskix kul'tur.

Yanko, T. (2017). Word order and accent placement in topics, foci, and markers of discourse continuity. In S. S. Alvestad (Ed.), *Comparative Slavic syntax and semantics. Oslo Studies in Language*, 9(1), pp. 45–57.

Zemskaja, E. A., Kitajgorodskaja, M. V., & Širjaev, E. N. (1981). *Russkaja razgovornaja reč'. Obščie voprosy. Slovoobrazovanie. Sintaksis* [Colloquial spoken Russian. General issues. Word formation. Syntax]. Moscow: Nauka.

Appendix A. Transcription conventions

Convention	**Meaning**
Separate lines in transcripts	Individual elementary discourse units (EDUs); boundary pauses
¦	Place of a potential segmentation into two (or more) EDUs
(0.23)	Silent pause and its duration, s
(ɥ 0.73)	Silent pause filled with a loud inhalation sound and its duration, s
(ə 0.20)	*uh*-like filled pause and its duration, s
(ɐ 0.33)	*ah*-like filled pause and its duration, s
(ɯ 0.48)	*um*-like filled pause and its duration, s
(ˀ 0.34)	Pause filled with glottal creak and its duration, s
(əɯ 0.62), etc.	Filled pauses of mixed nature
{laugh 1.02}	Laugh and its duration, s
{cl 0.12}, {st 0.23}, {gp 0.18}, etc.	Other non-verbal phenomena, such as a click of the tongue, snorting, gulping, etc.
/ \ – /\ etc. (placed before a word)	Pitch movements on stressed syllables of accented words
↑ ↓ →	Significant pitch movements on other syllables

Convention	Meaning
Underlining a word's stressed vowel	The given word bears the EDU's primary accent
Capitalization at the beginning of an EDU	Beginning of a new spoken sentence
.	Statement
?	Question
¿	Semi-statement
¡	Directive
@	Vocative
,	Default incompleteness
:	Incompleteness with further elucidation
…	Inexhaustiveness combined with an illocutionary completion
,,,	Inexhaustiveness combined with incompleteness
!	Exclamation
—	Splitting of one EDU into two or more parts as another EDU wedges in
( )	Inset
(*	"One-sided" inset
" "	Direct or semi-direct quotation
%	Co-construction in conversation
||	Mild internally-induced false start (the current EDU is not abandoned)
==	Severe internally-induced false start (the current EDU is abandoned)
⸾⸾	Mild externally-induced false start
≈≈	Severe externally-induced false start
~	Aposiopesis
=	Word truncation
$^{\text{ʔ}}$word$^{\text{ʔ}}$	Glottal stop at the word's onset/closure
$^{\text{ə}}$word$^{\text{ə}}$	*Schwa*-sound at the word's onset/closure
word$^{\text{ɯ}}$	Labial stop at the word's closure
word$^{\text{h}}$	Aspiration at the word's closure
a-a s-s ja-a j-ja	Phoneme lengthening
zvónit	Non-standard lexical stress
Italics	Accelerated tempo
Increased letter-spacing	Decelerated tempo
Grey	Perceptible phonetic reduction
Bold	Emphasis
Reduced font size	Heightened f0 register
Reduced font size $_{\text{below the baseline}}$	Lowered f0 register
#bum#	Onomatopoeias
<vot>	Presumable transcription of an uncertain fragment
<UNCLEAR: 2>	Unintelligible fragment: number of syllables

Appendix B. A fragment from the Russian Pear Chats and Stories corpus (Pears04), scores transcript

Line #	Pauses		TimeS	TimeE	Narrator		Commentator		Reteller	
0465	vp087	(0.22)	379.82	380.04						
0466			380.04	381.51					R-vE014	Togda /možno ešče \snačala? Could we start it over, then?
0467	vp088	(0.26)	381.51	381.77						
0468			381.77	382.27					R-vE015	\Značit, Well then,
0469			382.27	382.53					R-vN003	(ɥ 0.26)
0470	vp089	(0.14)	382.53	382.67						
0471			382.67		N-vE223	Skol'ko \↑ugodno. As much as you want.				
0472			383.08						R-vF002	(ə 0.65)
0473				383.67						
0474			383.67		*N-vL002*	{laugh 0.66}				
0475				383.73						
0476										
0477			384.31						R-vE016	(ɯ 0.10) čt= || (0.20) čt= == wh= … wh=
0478				384.34						
0479				385.02						
0480			385.02	388.64					R-vE017	vot (0.22) (ɯ 0.17) (əɯ 0.34) tam /mnogo dopustim \derev'ev s ėtimi grušami? are there, say, many trees with these pears?

Line #	Pauses		TimeS	TimeE	Narrator		Commentator		Reteller	
0481	vp090	(0.15)	388.64	388.79						
0482			388.79						R-vE018	Kogda /fermer vnačale ix /sobiraet, When the farmer collects them at the beginning,
0483			389.09	389.23			C-vN009	{hm 0.13}		
0484										
0485			390.05				C-vE019	–Po-okazano tol'ko \odno. Only one [tree] is shown.		
0486				390.63						
0487			390.63	390.81					R-vE019	s ≈≈ from …
0488				391.18						
0489	vp091	(0.29)	391.18	391.47						
0490			391.47	391.68	N-vE224	\Da, Yes,				
0491			391.68	391.90	N-vE225	/odno, one [tree],				
0492			391.90	393.10	N-vE226	dostatočno \bol'šoe. a rather big one.				
0493	vp092	(0.27)	393.10	393.37						
0494			393.37	393.74					R-vE020	\–A-a. Ah.
0495	vp093	(0.21)	393.74	393.95						

Line #	Pauses		TimeS	TimeE	Narrator		Commentator		Reteller	
0496			393.95						R-vE021	A o= || a on ne /ispol'zuet /→le-estnicu tam?… And … and doesn't he use a ladder or something?
0497			394.97				C-vE020	I || i /odet on kak-to ne \po-fermerski': And … and he is not dressed as a farmer:		
0498				395.86						
0499				397.15						
0500			397.15	399.07			C-vE021	ᵊv-v takie= || /–b-botinki u nego modnye,,, in … he's got fancy shoes,		
0501	vp094	(0.34)	399.07	399.41						
0502			399.41		*N-vL003*	{laugh 1.49}				
0503			399.50	399.75			*N-vL002*	{laugh 0.26}		
0504			399.75	400.31			C-vE022	/–brjuki,,, pants,		
0505										
0506			400.53				C-vE023	(tože \modnye,) also fancy,		
0507				400.91						
0508				401.32						
0509	vp095	(0.20)	401.32	401.52						

Line #	Pauses		TimeS	TimeE	Narrator		Commentator		Reteller	
0510			401.52				C-vE024	i samoe \interesnoe, and, most interestingly,		
0511			401.77						*R-vL004*	{laugh 0.50}
0512				402.22						
0513			402.22				C-vE025	čto-o (ˀ0.21) v ėtix modnyx /→brjukax on stoit na /**kolenke**, he is on his KNEE in these fancy pantsˀ,		
0514				402.27						
0515				404.95						
0516			404.95	406.41			C-vE026	*i posle ėtogo u nego takie* **pjatna**. and he's got these STAINS after that.		
0517	vp096	(0.37)	406.41	406.79						
0518			406.79	408.71			C-vE027	V obščem kakoj-to on ne poxož na \fermera. In short, he's like, he doesn't look like a farmer.		
0519	vp097	(0.44)	408.71	409.15						
0520			409.15						*R-vL005*	{laugh 0.64}
0521			409.42				C-vE028	Ne \odevajutsja (0.25) <tak> –↑fermery. Farmers don't dress like that.		
0522			409.60		N-vE227	Nu \/–da-a. Right.				
0523				409.79						
0524				410.34						
0525				411.22						

Line #	Pauses		TimeS	TimeE	Narrator		Commentator		Reteller	
0526	vp098	(0.12)	411.22	411.34						
0527			411.34						R-vE022	To est' on-n takoj /–tolstyj-j¿… So he is like fat?
0528			411.35	411.87	N-vF009	(ɐɯ 0.53)				
0529				413.11						
0530	vp099	(0.16)	413.11	413.27						
0531			413.27				C-vE029	V= || /–ˀ**us-sy** u nego,,, He's got a moustache,		
0532			413.28	413.86	N-vE228	\–Da-da-da! Oh yes!				
0533				414.22						
0534			414.22	415.09			C-vE030	/→bakenbardy… whiskers...		
0535			415.09				C-vN010	(ɥ 0.26)		
0536			415.21		N-vE229	\–Da-da-da! Oh yes!				
0537				415.35						
0538										
0539			415.58	415.93					R-vE023	–/Da? Is that right?
0540				415.96						
0541	vp100	(0.04)	415.96	416.04						

Line #	Pauses		TimeS	TimeE	Narrator		Commentator		Reteller	
0542			416.01						R-vE024	Usy /–bakenbardy… A moustache, whiskers…
0543			416.60		N-vE230	(ˀ 0.19) Po /\lestnice! Ladder!				
0544			416.73				C-vE031	Krasnaja <ta=> ≈≈ A red …		
0545				417.29						
0546				417.41						
0547				417.46						
0548			417.46	418.57	N-vE231	On \podnimaetsja po /lestnice, He climbs a ladder,				
0549			418.57		N-vE232	\spuskaetsja ≈≈ [then] comes down …				
0550			418.70						R-vE025	On ne /lysyj?! Is he bald?!
0551				419.55						
0552				419.70						
0553	vp101	(0.17)	419.70	419.87						
0554			419.87	420.31			C-vE032	\Net, No,		
0555			420.31				C-vE033	u n'= || \/↑kudrjavyj. he … he's got curly hair.		
0556			421.04						R-vE026	\Aga. OK.
0557				421.27						
0558				421.31						
0559	vp102	(0.18)	421.31	421.49						

CHAPTER 2

The basic unit of spoken language and the interfaces between prosody, discourse and syntax

A view from spontaneous spoken Hebrew

Shlomo Izre'el
Tel Aviv University

Looking at spoken language as an integrative whole, where prosody, syntax and discourse features interplay as to conveying information, I will try to figure out the best methodology for its research by advocating that the best candidate to be regarded as the basic unit of spoken discourse is the *utterance*. Arguments brought will be mainly phonetic, phonological (prosodic), informational, and syntactic. In addition, arguments from pragmatics and conversation analysis will be mentioned.

Keywords: prosodic units, utterance, segmentation, syntax, clause, predicate, spoken language, Hebrew

1. Introduction

When people are engaged in oral communication, speakers transmit information linearly, manipulating the segmental stretch by prosodic indications of its structure. Looking at spoken language as an integrative whole, where prosody, syntax and discourse features interplay as to conveying information, I will try to figure out the best methodology for its research by advocating that the *utterance* (or *information set*) is the best candidate to be regarded as the basic unit of spoken discourse. The approach taken here is built on the premise that syntax, information structure, and prosody integrate in spoken language structure, forming a coherent unity.

The corpus used for this research contains the major part of the original, reliably-transcribed recordings of *The Corpus of Spoken Israeli Hebrew (CoSIH)*, recorded between August 2000 and October 2002 (see *CoSIH* 2012 webpages at <cosih.com/english/index.html>). Preliminary research has revealed that text types may differ in their segmentational strategies. Therefore, only texts which include

https://doi.org/10.1075/scl.94.02izr
Additional files available at http://doi.org/10.1075/scl.94.02izr/audio

mostly spontaneous conversations have been taken for this research, which conforms well with the aims of this volume. The analyzed part consists of approximately 4.45 hours of spontaneous daily conversations (= ca. 37,000 words in Hebrew orthography), recorded by 36 volunteers.[1] The total number of identified speakers is 137. In its totality, the analyzed corpus thus consists of 8,391 *utterances* (Utts) or 13,730 *information modules* (IMs), including fragmentary or undeciphered IMs (and Utts).

2. Units of spoken language: Definitions and terminology

It is commonly accepted, that prosodic units encapsulate corresponding segmental units, together constituting discourse units (cf. Cruttenden, 1997, p. 7; Féry, 2017, pp. 36–37; Kibrik, Korotaev, & Podlesskaya, this volume, Part I; Szczepek Reed, 2011, Section 3.2.1; among many others).[2] Discourse units in themselves can either overlap or interface with syntactic units. Indeed, there is broad consensus that prosodic units encapsulate coherent structural, functional segmental units. While the hierarchy of prosodic units seems more or less established and agreed upon by most scholars (Selkirk, 2001, p. 896), there is an ongoing debate on the type of unit that will serve as the unit of reference for the study of spoken language, notably its spontaneous, daily conversational varieties. As our concern here is the search for units of reference for the study of spoken, more specifically: spontaneous discourse, I shall focus on those units in two of the higher hierarchical levels where all three components interface:

a. Level 1: *prosodic module* (PM), *segmental module* (SM) and *information module* (IM);
b. Level 2: *prosodic set* (PS) and *utterance* (Utt) or *information set*.[3]

[1] In Hebrew standard orthography, single-consonantal function words and enclitic pronouns are written bound together with their host. This characteristic of Hebrew orthography, among other features of the language, makes the number of words much reduced than an equivalent corpus in European languages. The ratio between the number of words in Hebrew vs. European orthographies can be estimated at about 2/3.

[2] Some prosodic modules encapsulate semantically-void segments (e.g., *e* ‘uh’; *m* / ‘what?’), so that information (mostly regulatory; see 3.4) is conveyed only by prosody.

[3] There are two possible higher-level units of reference that may be discerned. The highest will possibly be the *period* (cf. Izre’el & Mettouchi, 2015, Section 2.4). However, this unit, if proven valid, would probably not include syntax (= sentence structure) at its interface. An intermediate level between the two may well show interface features between prosody, information structure and syntax, at least partially.

2.1 Prosodic and information (discourse) units

2.1.1 *Prosodic module (PM), segmental module (SM) and information module (IM)*

A *prosodic module* (henceforth: PM) is the smallest prosodic unit that can be perceived by prosodic contours and prosodic boundaries. It can thus be regarded as the first-level unit of prosody relevant for the study of spoken discourse. The PM encapsulates a segmental unit of language to be termed *segmental module* (SM), forming together an *information module* (IM). The boundaries of either a SM or an IM are therefore defined by prosody. As we shall see below, there are two main classes of boundaries: major (indicating terminality) or minor (indicating continuity). Both are indicated by their respective boundary tones. A major boundary is also the boundary of a *prosodic set* (see Section 2.1.2).

The unit which has been termed here *prosodic module* has been known in the research literature in many related terms, among which a widely used one is *intonation unit* (see, among many others, Chafe, 1994). The term used here, *prosodic module*, seems to me preferred over terms using *intonation* (or *tone*) rather than *prosody*, since intonation is more restricted in scope than *prosody* (Crystal, 2008, s.v.). As for the term *unit* (or *group* or *phrase*), it is too general, whereas *module* suggests the capacity of a unit to be used either independently or in combination with similar units.

PM is usually regarded – sometimes even defined – as consisting of a "single coherent intonation contour" (Du Bois, Cumming, Schuetze-Coburn, & Paolino, 1992, p. 17). However, a coherent intonation contour, while rather easily perceivable, is hard to define in itself by acoustic, formal terms, nor is it easy to define a PM by any other internal criteria alone. In practice, segmentation of a discourse flow into PMs is made chiefly by detecting their boundaries, whereas internal criteria are brought into consideration only secondarily (e.g., Cruttenden, 1997, Section 3.2).

Segmentation into PMs in *CoSIH* was carried out applying both external and internal criteria, that is, by detecting boundaries of PMs and by looking at the internal structure of the pitch contour. In accordance with previous research on various languages, we have found valid the following four major perceptual and acoustic cues for boundary recognition: (1) final lengthening, (2) initial rush, (3) pitch reset, (4) pause (Amir, Silber-Varod, & Izre'el, 2004). The internal criteria used – apart from an impressionistic-perceptual conception of a contour, were: (1) declination (Cruttenden, 1997, Sections 4.4.4.4, 5.5.1; Wichmann, 2000, Section 5.1.1), (2) isotony (Izre'el & Mettouchi, 2015, pp. 23–25, following Du Bois, 2006; Wichmann, 2000, Section 4.3).

It should be noted that none of the four cues for prosodic boundaries is in itself a necessary or sufficient cue for the existence of a PM boundary, and languages may differ in their most prominent cue for delimitation of PMs (Hirst & Di Cristo, 1998, passim; Izre'el & Mettouchi, 2015, pp. 24–25). Research on *CoSIH* data has shown that final lengthening is the highest in hierarchy among acoustic features presented at a PM boundary, whereas initial rush occupies the last position in this hierarchy (Amir et al., 2004). A different approach to the same data will consider tempo change as the highest in hierarchy and pause as the lowest.[4]

Obviously, there can be no sharp division according to pitch levels of final syllables, upon which the discourse function of the module boundary can be perceived. Nevertheless, scholars usually agree on binary functional categories for discourse analysis, which will be termed here major and minor. Major boundaries signal finality; minor boundaries signal continuity. Both categories have variants. Tone-variants of minor boundaries are less relevant to our discussion. The default major boundary is usually signaled by a falling tone. The other major boundary indicates what Du Bois et al. called "appeal", that is, "seeking validation response from listener" (Du Bois et al., 1992, Section 6.3, 1993, Section 3.3), usually indicating the final tone of polarity (yes/no) questions. This boundary is indicated by a (usually high) rising tone at the end of the PM, which may also show the end of a graduate, longer upward movement. The practice of indicating both major boundaries has been applied in *CoSIH* (Izre'el, 2002, following Du Bois et al., 1992, 1993).

The falling tone at the end of a PM seems to be a natural consequence of the respiratory mechanism (Lieberman & Blumstein, 1988, pp. 198–203, essentially following Lieberman, 1967, Chapter 5). Due to this natural basis of the PM, Lieberman has suggested to term it "breath-group", noting that pausing for inspiration between two adjacent breath-groups is not a necessary condition. If a breath-group terminates in a falling pitch, it is accordingly viewed as unmarked. In contrast, any breath-group (=PM) that does not end in a fall, is seen as marked.

As noted by many, the non-falling (rising or level) final tone usually implies non-finality or continuity (see, inter alia, Chafe, 1994, p. 140; Hirst & Di Cristo, 1998, p. 27; Lieberman, 1967, p. 109). Bolinger (1972, p. 28, 1986, Chapter 9) takes non-finality to be a universal criterion entailing rising or high pitch on both statements and questions (wording by Hirst & Di Cristo, 1998, p. 27; but cf. Lieberman, 1967, Chapter 6, for some reservations; Hirst & Di Cristo, 1998, p. 1, Section 2.2). For Brazil (1995, Chapters 3, 16, 17), speech increments can be categorized by a binary opposition, according to their final tone: proclaiming (final falling tone) and referring (final rising tone), where the latter increments would include both increments to be continued by the same speaker or by an interlocutor, namely, certain types of questions. From our point of departure of seeking the basic discourse unit in

[4] This finding was endorsed later also by the Hebrew part of CorpAfroAs – The Corpus of AfroAsiatic Languages (Izre'el & Mettouchi, 2015, p. 23).

spontaneous, conversational speech, we look at communicative signals. For this goal, it seems preferable to draw the boundary between signals for completeness or finality versus signals for incompleteness or continuity as transmitted to the interlocutor.

The formal indications of the various boundaries and their functional status in discourse are summarized in Table 1. Note that the table lists default indications of boundary tones. Taking the above analysis into account, it can be claimed that a unit ending in a major boundary will be regarded as basic.

Table 1. Default indications of boundary tones

	MINOR			MAJOR
TONE	fall	high rise	other	level/rise
MARKEDNESS	unmarked	marked		
INDICATION	finality			continuity
SPEAKER SWITCH	enabled	required	enabled	constrained

Finally, fragmentary PMs should be mentioned. These are indicated by acoustic features such as incomplete or incoherent pitch contour, abrupt final syllable or final glottal stop (Du Bois et al., 1992, Section 4.4). At times, the end of a fragmentary PM correlates with a fragmentary word.

A PM encapsulates a *segmental module* (SM). The combined prosodic-segmental modules form an *information module* (IM). The PM is regarded as the first-level unit of prosody relevant for discourse analysis. Accordingly, an IM will be viewed as the first-level discourse unit.

As PM boundaries are indicated by prosodic signifiers, they can also indicate IM boundaries. Therefore, we shall distinguish between major IMs and minor IMs, defined according to their respective PM boundaries.

2.1.2 *Prosodic set (PS) and utterance (Utt) or information set*

A *prosodic set* (PS) is the second-level unit of prosody relevant for the study of spoken discourse. PS is defined as one or more PMs of which the last PM ends in a major boundary (signaling finality), where any (optional) previous PM carries a minor boundary tone (signaling continuity).[5] The mathematical concept of *set*, which can consist of any number of members, including a single one, has been adopted here, since a *set* can consist of either a single *module* or more.[6] Thus, an *utterance*

[5] A few self-standing PMs/IMs cannot be defined by tonal features, as they consist of voiceless paralinguistic elements like clicks (indicating negation) or hush sounds [ʃ(ː)]. The status as Utt of each such unit will be approved by their position within the discourse sequence.

[6] I thank Alexander Sodin for his clarifications on the mathematical concept.

(Utt) is the discourse unit that can be defined as an *information set*, consisting of one or more IMs. As such, it consists of one or more SMs encapsulated by their respective PMs. The boundary of an Utt will thus be – in its default manifestation – a major prosodic boundary.

It may be useful to recall at this juncture that a sign of completeness, namely, a major boundary, need not indicate the end of a turn but the end of a *prosodic set* and therefore of an *utterance*, as we shall see in Example (1) below. This may be compared, *mutatis mutandis*, to what has been termed by conversation analysts "Transition Relevant Place" (TRP), noting that it need not be an absolute sign for turn ending but a potential completion point (Ford & Thompson, 1996; cf. Liddicoat, 2007, pp. 57–63; Sacks, Schegloff, & Jefferson, 1974).

Example (1) and Figure 1 illustrate a stretch of speech consisting of seven IMs constituting four Utts.[7] The speaker describes a Mongolian castle with its surroundings.

(1) [1] naˈgid hatiˈra hi meruˈbaat / (0.304)
'Say the castle is square?'
[2] hi riˈbua / (0.505)
'It is a square?'
[3] kiloˈmeter | (0.263)
'A kilometer'
[4] miˈkol eː |
'from each uh'
[5] piˈna |
'corner'
[6] jeʃ ʦav ||
'there is a turtle.'
[7] ʃe ʃoˈmer ||
'That guards.' [source: OCh_sp1_128–134]

[7] 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. References follow the system used in *CoSIH*; speakers are referred to as sp1, sp2, and so forth.

Notation: | minor boundary; || major boundary; / major boundary with "appeal" tone; - fragmentary (truncated) module; — truncated word; (0.234) pauses (measures in seconds).

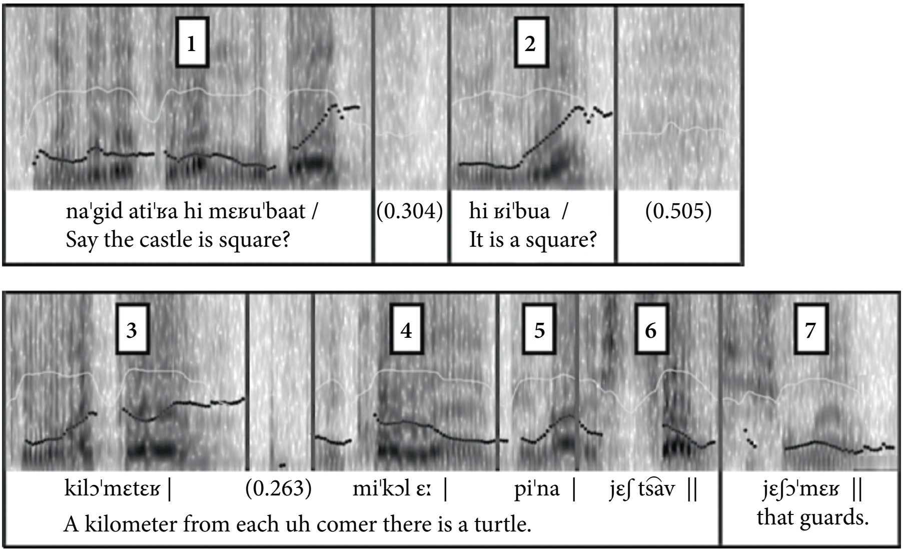


Figure 1. A stretch of speech consisting of 3 IMs constituting 3 Utts

IMs [1], [2], [6] and [7] are major IMs. PMs [6] and [7] end in a fall, implying finality. PMs [1] and [2] end in a rise, also implying finality, but interpreted as "appeal". In these case, the speaker does not forward the turn to his interlocutor, because the questions are part of his own knowledge, transmitted in this way to the recipient, thus building a common ground for what will come next (Warren, 2016, Section 3.3).

IMs [3], [4], [5] are minor IMs. PM [3] and [4] end in a level tone; PM [5] end in a rise. It will be noted, that lengthened level-tone syllables correlate with syntactic proximity (Silber-Varod, 2011), which is the case also for IMs [3] and [4] here, expressing together a noun phrase with an adjunct: *kiloˈmeter miˈkol piˈna* 'A kilometer from each corner'. This noun phrase functions as a topicalized locative adverbial phrase for the existential clause *jeʃ ʦav* 'there is a turtle' that follows.

In terms of Utts, IMs [1], [2], [7] constitute each an Utt on its own, whereas IMs [3], [4], [5] [6] constitute a single Utt: The first three IMs have minor boundaries; IM [6] ends the series with a major boundary. The last Utt, IM [7], follows the Utt expressed in IMs [3]–[6], continuing the line of thought. For the syntactic implications of this sequence see Section 3.3 below.

2.2 Syntax: The clause

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.

Like many recent approaches to clause structure, I take the predicate to be its core component. However, in contrast to common views, 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, *clause* is defined as a syntactic unit consisting minimally of a predicate. The definition of *clause* is thus dependent on the definition of *predicate*, which will be brought forward below. Clauses can be unipartite or bipartite. Unipartite clauses consist of only a predicate component; bipartite clauses consist of both a predicate component and a subject component.

Most significantly, in Hebrew, any part of speech can function as predicate (or a predicative nucleus). Some simple cases are illustrated in Example (2) (noun), Example (3) (prepositional phrase) and Example (4) (predicate domain with a nominal nucleus).[8]

(2) ha='kol **ʃvi'l-im** ||
DEF=all path-PL
'All are dust-roads.' [source: OCh_sp1_192]

(3) a'ni **be='kurs** ||
I in=course
'I am taking a course.' [source: OCD_3_sp1_060]

(4) ze **ha=ba'sis le=χol=da'var** ||
DEM.SGM DEF=basis to=all=thing
'This is the basis for everything.' [source: P931_1_sp2_044]

Larger complexes can function as predicates as well:

[8] Predicates (or predicate domains) are marked by boldface characters. The notion of predicate domain may seem *prima facie* equivalent to the notion of predicate phrase (or, rather, verb phrase) as commonly used in other schools of thought (e.g., Chomsky, 1957; and followers). However, as already noted by Chomsky (1965), "[f]unctional notions like 'Subject', 'Predicate' are to be sharply distinguished from categorial notions such as 'Noun Phrase', 'Verb', a distinction that is not to be obscured by the occasional use of the same term for notions of both kinds" (p. 68). Furthermore, the term *phrase* seems to be contradictory to the notion of a complete sentence or clause (cf. e.g., Harris, 1951, p. 14), which in the framework used here will make a false claim as regards the very notion of sentence or clause.

(5) ze **ma ʃe a'marti=l=a** ||
DEM.SGM what that I.said=to=her
'This is what I said to her.' [source: OCD_1_sp2_009]

A verb is not a primary predicate, as it includes a pronominal subject and thus constitutes a clause on its own:

(6) **χa'ʃav**-ti |
thought-I
'I thought,' [source: Y32_sp1_087]

A verb like *χaʃavti* 'I thought' in (6) includes a predicate in the form of a stem /*χaʃav-*/ 'thought' and a bound pronominal 1SG subject /*-ti*/. Still, verbs can function as a clausal predicate when an appositional subject is added, as in (7):

(7) ani **χa'ʃav-ti** ||
I thought-I
'I thought.'[9] [source: C711_3_sp2_028]

Thus, verbs are always bipartite clauses, the same as are the clauses in Examples (2)–(5) above. A unipartite clause is illustrated in (6). Sp1 had told Sp2 about a ride he made in Mongolia on a local breed of horses, and Sp2 was suggesting that they were mules rather than horses. Sp1 insists that this kind of animal is a genuine horse, and Sp2 responds by a verifying question:

(8) [1] sp2: sus ma'maʃ /
horse real
'(Is it) a real horse?'
[2] sp1: sus sus |
horse horse
'(It is) a real horse,'
[3] rak jo'ter na'muχ ||
only more short
'but shorter.'
[4] rag'laim mekutsa'rot ka'ele ||
legs shortened sort.of
'(It has) sort of shortened legs.'
[source: OCh_sp2_091; sp1_286–288]

[9] Depending on speaker and on register, the added pronoun may or may not carry pragmatic information. Note: This utterance is hard to decipher in the recording due to overlap.

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. In other words, all clauses are unipartite, consisting of only predicate domains (Izre'el, 2018a, 2018b).

Since any part of speech can function as predicate and since the predicate need not be related to an argument, or, more specifically, it need not be seen as depending on a subject – a new perspective of what consists of a predicate is in order. As mentioned, a discourse-related approach is taken.

Thus, the *predicate* (or the predicate domain) is viewed as the component carrying an individual piece of information within the discourse context, which by default will include a newly introduced element (cf. Chafe, 1994, p. 108). As such, a predicate may be seen as the default representation of the comment (Hockett, 1958, p. 201; Lyons, 1968, Section 8.1.2; Sornicola, 2006). The predicate (or the predicate domain) carries the modality of the clause (Izre'el, 2012, 2018a, 2018b). By default, the focus of the clause will be found within the predicate domain.

3. In search of the basic unit of spoken language

3.1 The interface between prosodic, information (discourse), and syntactic units

It will be recalled (see Section 1), that the approach taken here is built on the premise that syntax, information structure and prosody integrate in spoken language structure, forming a coherent unity. It has been claimed – and now has become almost a consensus – that the default structural form of an IM is the clause; in other words, the default domain of the clause is the IM (Chafe, 1994, pp. 65–66; Halliday, 2014, Section 1.2.2; Kibrik & Podlesskaya, 2009; Kibrik et al., this volume, Part I; to name only a few). However, published statistics may not support such claims. Table 2 lists the percentage of overlap between clauses and IMs in different languages.

The overall impression from these data is that only about half or less of the IMs do include clauses (depending on the language, the studied corpus, and the theoretical approach taken).[10] As seen, in Hebrew the count came up with less than half.

[10] The corpora upon which these data have been extracted are limited in scope; not all attest to everyday conversation language; the represented languages represent diverse structural systems; theoretical approaches are different; and yet, the overall picture seems indicative.

Table 2. Ratio between IMs and clauses in several languages

English	54%	Iwasaki & Tao, 1993, p. 3
	60% (substantive units)*	Chafe, 1994, pp. 65–66
	48%	Croft, 1995, p. 849
Japanese	42% ~ 45%	Iwasaki, 1993, p. 41; Iwasaki & Tao, 1993, p. 3
	68%	Matsumoto, 2003, p. 58
	50% ~ 68%	Den et al., 2010
Russian	70%	Kibrik & Podlesskaya, 2006, Section 8
Mandarin	40% ~ 47%	Iwasaki & Tao, 1993, p. 3; Tao, 1996, p. 72
Wardaman	50%	Croft, 2007, pp. 11–12
Sasak	32% ~ 52%	Wouk, 2008, pp. 150, 158
Hebrew	42% ~ 47%	Izre'el, 2005, Section 6.1

* For substantive vs. regulatory units see Section 3.4.

A more recent quantitative evaluation of *CoSIH* reveals the following data: 45% of all IMs are Utt singletons (i.e., constituting an Utt in themselves); more than 90% of the syntactically-relevant singletons (i.e., non-fragmentary, complete IM=Utt that are syntactically analyzable) consist of a single clause each. These Utts may thus conform to the hypothesis that an IM is the primary domain of the clause. Still, 18.6% of all clauses spread over more than a single IM, many of them without apparent grammatical motivation, that is, prosodic boundaries do not coincide with grammatical ones, or they do not show any pragmatic motivations. In many of these cases, cognitive or interactional motivation seem to generate prosodic boundaries (Shor, 2016).

Having these data at hand, it seems that the wide consensus that the IM is the default domain of the clause needs reevaluation. Furthermore, 2/3 of the corpus consist of other configurations that need to be accounted for. Notably, many singletons (1/3) or most of the multi-IM Utts, as well as some intra-Utt IMs, consist of other configurations than a single, complete clause.

The following two excerpts illustrate instances where the relationship between IMs and clauses are incompatible with the common hypothesis:

(9) [1] tiv'dok et=ha=ktsi'tsot hem b=a= |
check ACC=DEF=meat.balls they in=DEF=
'Check the meatballs. They are in the'
[2] 'frizer ||
freezer
'freezer.' [source: C711_1_sp3_001–002]

(10)	kiloˈmeter |	(0.263)	miˈkol	eː |	piˈna |	jeʃ	ʦav ||
	kilometer		from.all	uh	corner	EXT	turtle

'A kilometer from each uh corner there is a turtle.'

[source: OCh_sp1_130–133]

In (9), IM [1] consists of a full clause and the beginning of another clause; IM [2] consists of the predicate of the bipartite clause that started in IM [1]. In (10), already presented above as IMs [3] to [6] of Example (1), four IMs constitute a single clause, with no structural or pragmatic reasons for phrase distribution among IMs.[11] In fact, there are two basic options for setting out the interface between prosodic or information units and syntactic units, focusing on the clause:

1. In conformity with the consensus, the IM will be regarded as the default domain of a clause, notwithstanding accountable exceptions.
2. In contrast, the Utt will be regarded as the default domain of the clause, notwithstanding accountable exceptions.

As the consensual view that the domain of the clause is the IM seems to be flawed by numerous unaccountable exceptions, the second option should be checked out.

3.2 Hypotheses

Given the arguments above, the following hypotheses can be put forward for evaluation:

a. The *utterance* (Utt) is the default domain of the clause.
b. By default, an Utt will consist of a single clause.
c. The Utt is the biggest information unit that can contain a clause. A clause cannot spread beyond the boundaries of a single Utt. In other words, a major prosodic boundary indicates the terminal boundary of a clause. Any subsequent stretch will therefore be the beginning of a new clause.

In the following sections (Sections 3.3–3.5), these hypotheses will be tested. Section 3.6 will bring forward some interim conclusions, further addressing the term *sentence* with regard to spoken language. Apparent exceptions will then be reviewed (Section 3.7).

[11] For cognitive and interactional motivations for prosodic phrasing, see Shor (2016).

3.3 One single utterance consists of a single clause

As mentioned above (Section 3.1), 45% of all IMs are Utt singletons (i.e., constituting an Utt in themselves); 90.7% of the syntactically-relevant singletons (which is about a third of all syntactically-relevant Utts) consist of a single clause each. There are numerous such instances in the examples cited above (e.g., Example (1), IM [2]; Examples (2)–(5); and others). These Utts thus conform to either the hypothesis that an IM is the primary domain of the clause or to a hypothesis that an Utt is the primary domain of the clause.

It has also been mentioned, that 18.6% of all clauses spread over more than a single IM, many of them without apparent grammatical or pragmatic motivation. Examples (9) and Example (10) above are two instances of such Utts. Thus, the view that an Utt rather than an IM would be the domain of a clause seems preferable, as it covers also these cases.

There are, however, cases where a segmental unit seems not to consist of a full clause. The question now arises what is a clause. As defined in Section 2.2, *clause* is a syntactic unit consisting minimally of a predicate, whereas a *predicate* (or the *predicate domain*) has been viewed as the component carrying an individual piece of information within the discourse context, which by default will include a newly introduced element. As such, a predicate may be seen as the default representation of the comment. The predicate (or the predicate domain) carries the modality of the clause. By default, the focus of the clause will be found within the predicate domain. As mentioned, clauses can be bipartite, consisting of both a predicate component and a subject component, or unipartite, consisting of only a predicate component. One stretch of unipartite clauses, relatively straightforward, has been displayed in (8), where clauses are encapsulated either by Utts (IMs [1], [4]) or by IMs constituting together a single Utt (IMs [2], [3]). Another, more obscure case has been cited above in IMs [3] to [7] of Example (1), repeated here as Example (11):

(11) kilo'meter | (0.263) mi'kol eː| pi'na | jeʃ ʦav ||
kilometer from.all uh corner EXT turtle
'A kilometer from each uh corner there is a turtle.'
ʃe ʃo'mer ||
that guard.PTCP[SGM]
'That guards.' [source: OCh_sp1_130-134]

The second Utt includes what would usually be defined as an "afterthought", consisting of a relative clause attributive to *ʦav* 'turtle', the final component of the previous Utt. Semantically, the SM which constitutes the second Utt is indeed related to *ʦav* 'turtle'. From the syntactic point of view, it certainly accords with all characteristics that define *predicate*, and therefore a complete unipartite clause: It carries new information, pronounced in assertive (or indicative) modality, and carries focus; the two latter features being signaled by prosody: Declarative modality is indicated by the intonation contour (Debaisieux & Martin, this volume, Part I; cf. Martin, 2015, pp. 69–71) and focus both by the independent PM and by the marked stress (Section 2.1.1, Figure 1). For further observations see Izre’el (2018a, 2018b). Similar cases of syntactic relations between clauses (or sentences; cf. Section 3.6) have been termed *insubordination*, defined as “the conventionalized main clause use of what, on prima facie grounds, appear to be formally subordinate clauses” (Evans, 2007, p. 367; see further Bossaglia, Mello, & Raso, this volume, Part I; Debaisieux, 2013; Mithun, 2008; for Hebrew *ʃe* ‘that’, see Inbar, 2016).

3.4 Utterances with no syntactic contents

Speech stretches can carry either substantial or regulatory information. Referring to what he termed *intonation units* (our IMs), Chafe (1993, 1994, Chapter 5) distinguishes between two basic types of units – *substantive* and *regulatory*:

> Substantive intonation units are the contentful stretches of speech that include ideas of people, objects, events and states. They are in a sense what language is about…. Regulatory intonation units are those whose primary function is, in one way or another, to regulate the flow of information. (Chafe, 1993, p. 37)

For Chafe (1994), “[r]egulatory intonation units coincide to a large extent with the devices that have been discussed under the label *discourse markers*” (p. 64). Example (12) illustrates an Utt consisting of two regulatory IMs, uttered as a backchannel.

(12) a | okej ||
Oh okay
‘Oh, okay.’ [source: C842_sp1_154–155]

It will be noted that many regulatory units are syntactically analyzable, and will thus form part of the data for the interface between prosody, information and syntax. Example (13) illustrates this type of regulatory Utts. This excerpt is taken from the final part of a telephone conversation in which only one of the interlocutors in heard. The speaker is talking to a person who is driving during rush hour. After more than 12 minutes of speaking, the speaker is trying to end the conversation, and we can hear the following units of speech, meant to indicate precisely the wish to end the conversation:

(13) tov | (0.304) jakirati | (0.117) tamʃiχi laχ bapkak | (1.328) tamʃiχi
good my.dear continue to.you in.the.traffic.jam continue
laχ bapkak ||
to.you in.the.traffic.jam
'Well, my dear, keep to your traffic jam, keep to your traffic jam.'
[source: C514_2_sp1_305–308]

3.5 Expanded configurations

Utts can contain more than just a single clause: a clause with an additional non-clausal elements (Section 3.5.1) or two or more clauses, with or without non-clausal elements (Section 3.5.2).

3.5.1 *Clause+*

"*Clause+*" is used here as a label for Utts consisting of a single clause plus a non-clausal component. Such components can be regulatory components, either forming a separate IM as in (14) or joining a clause within a single IM (15);[12] repeats (16), repairs (17), among others.

(14) ma | maˈraχt=otam /
what you(SGF).spread=them
'What? Did you spread them?' [source: C714_sp1_015–016]

(15) naˈgid ha=tiˈra hi meruˈbaat /
say DEF=castle she square
'Say the castle is square?' [source: OCh_sp1_128]

(16) ʃviˈleː= | ʃvile=aˈfar ||
roads.of= roads.of=dirt
'Roads… dirt roads.' [source: OCh_sp1_181–182]

(17) aˈval hu=loː | aˈni lo roˈa=oto po beχˈlal ||
but he=NEG I NEG see=him here at.all
'But he is not … I do not see him here at all.'
[source: C711_4_sp3_020–021]

[12] Example (1), IM [1], repeated here as (15).

3.5.2 *Clause clusters*

The phrase *clause cluster* is used here to convey a series of two or more clauses brought together, conveying a single, integrated message. Occurring together within a single Utt helps in communicating their informational unity. Examples (18) and (19) illustrate single Utts consisting of two IMs including a single clause each.[13]

(18) im ha'ju miʃtam'ʃim kol jom | ‡ az ejn baa'ja ||
if they.were using every day ‡ then NEG.EXT problem
'If they would use (it) every day, then there would be no problem.'
[source: C612_3_sp2_045–046]

(19) ha'jiti no'tenet=leχa et=ha=mafte'χot ʃel=ha='bait=ʃeli ‡ ha'jita
I.was give=to.you ACC=DEF=keys of=DEF=house=mine ‡ you(SGM)
ja'ʃen ets'li ||
sleep at.me
'I would have given you the keys for my home, (so that) you would have slept over at my place.'
[source: OCD_3_sp2_057]

In (18), a correlative structure (if… then…) indicates the relationship between the two clauses. Prosodic packaging in a single Utt further suggests the relationship between the two clauses. In contrast, Example (19) illustrates a case where a single Utt consisting of a single IM includes two clauses. There are no segmental markers like prepositions or other particles to indicate the dependency of the message conveyed by the second clause upon the one conveyed in the first clause. The relationship between the two clauses is established by their prosodic packaging in a single Utt, all the more so as they come together within a single IM. Intermediate cases are also possible; one such case is illustrated in (20), where only the first clause is marked segmentally. Similar to the case illustrated in (19), the Utt in (21) consists of a single IM.

(20) im hi=tir'tse laa'vor=iti | ‡ sa'baba ||
if she=she.will.want to.pass=with.me ‡ fine
'If she wants to move with me – (then it's) fine.'
[source: Y32_sp2_200–201]

(21) ma at ro'tsa ‡ ʃe hu=ja'gid=laχ ‡ 'gili 'χara ‡
what you.SGF want ‡ that he=will.say=you ‡ Gili shit ‡
at sa'baba /
you.SGF fine
'What do you want him to tell you: "Gili is shit; you are okay?"'
[source: OCD_3_sp1_024]

[13] Clause boundaries are indicated by ‡.

In this case, there is one particle indicating the syntactic relationship between the first and the second clause, whereas the following two clauses convey a single hypothetical quotation following an introductory direct speech presentational verb: *jagid* ‘he will say’. Closer prosodic packaging with correlative structures are also attested, where a correlative structure is attested within a single IM, as in (22). An Utt consisting of more than two IMs is illustrated in (23a).

(22) aval im hi toˈva ‡ az keˈdaj |
but if she good ‡ then worthy
‘But if she is good, (then it is) worthy,’ [source: Y33_sp2_111]

(23) a. ani ˈkodem ˈsama |‡ (0.619) m̩kavˈtʃetʃet niˈjar | ‡ keˈdej ʃ
I first put ‡ crumple paper ‡ so.that that
jihˈje leˈze | ˈroχav | (0.689) ˈomek kiˈilu ‡ aˈni | (0.294)
it.will.be to.this width depth like ‡ I
madbiˈka ˈkodem | (0.463) kivtʃuˈtʃim ʃel=niˈar kiˈilu / ‡ ve
glue first crumples of=paper like and
ˈal=ze oˈsa ||
on=this do
‘I first put … crumple paper, so that it gets width, that is depth; I first glue like paper crumples, and on this I make (it).’
[source: C714_sp4_015-022]

Example (23a) exhibits an Utt consisting of eight IMs. There are six clauses in this Utt, including one that makes a repair to the last component in the previous clause (repairing *roχav* ‘width’ by *omek* ‘depth’). This example further illustrates one of the rare cases (23b) where a single IM includes – in addition to its own clausal component – also the beginning of a new clause:

(23) b. ˈomek kiˈilu ‡ aˈni | (0.294) madbiˈka ˈkodem |
depth like ‡ I glue first
‘that is depth; I first glue,’ [source: C714_sp4_019–020]

3.6 Interim conclusions

We have seen (Section 3.3) that there is a non-negligible number of clauses that are not confined within the boundaries of a single IM. On the other hand, probably all occurrences of clauses that – traditionally – seem to overflow a single Utt can be accounted for, with the definition of a clause used here (Section 3.3, Example (11)). We have further seen that we can account for cases where an Utt includes more than a single clause (Section 3.5). In fact, about 2/3 of the syntactically-relevant Utts consist of more than a single clause. Therefore, we can add the following insights as regards the interface of Utts with syntactic units:

a. An Utt can include non-clausal components in addition to a clause (*clause*+) or a clause cluster.
b. An IM forming part of an Utt consisting of more than a single IM will include either a full clause or a phrase.
c. Less frequently, an IM forming part of an Utt that consists of more than a single IM can include more than a single clause.

The interface between prosodic and segmental units is presented in Table 3.

Table 3. The interface between prosodic and segmental units

Prosodic units	*Discourse units*	*Syntactic units*
Prosodic set (PS)	Utterance (Utt)	Clause / Clause+ / Clause cluster
Prosodic module (PM) (one of two or more in a PS)	Information module (IM) (one of two or more in an Utt)	Phrase / Clause (/ Clause+ / Clause cluster)

Adding these arguments to the basic phonetic argument for the structure of an Utt (Section 2.1.1), I believe that we have a strong case to claim that the basic unit of spoken language is the Utt than claiming the same for the IM.

For Kibrik and Podlesskaya (2008),

> sentences are groups of EDUs [Elementary Discourse Units][14] found not only in written language but also in speech…. Overall, sentence is a difficult, non-elementary, and elusive notion. Unlike clause and EDU, sentence should not be considered a basic unit of language. (Kibrik & Podlesskaya, 2008, p. 85)

For Halliday (2014), a "clause complex realizes a semantic sequence of projection or expansion; and it is, in turn, realized by a sequence of tones in speech and by a sentence in writing" (p. 435).

Very much like the view of the Utt as an *information set* and its defining equivalent prosodic unit as *prosodic set*, one can think of the syntactic unit enclosed by the Utt as a *clause set*. Thus, we include Utts consisting of a single clause, a clause+, or a clause cluster (or complex, if one adopts Halliday's term). Along Halliday's line of thinking, we can plainly see the *clause set* as a spoken *sentence*.

Indeed, as put forward by Kibrik and Podlesskaya, in terms of syntax, sentence should not be regarded as a basic unit of language. Surely, it is clause that is more basic than sentence. However, when looking at spoken discourse structure, it is the Utt that should be regarded as basic, when one analyzes both its information capacities and its interface with syntax.

[14] An EDU is prosodically equivalent to a PM, but defined according to multiple criteria, including syntax, as an EDU "coincides with a clause" (Kibrik & Podlesskaya, 2006, 2009, Chapter 4; see also Kibrik et al., this volume, Part I).

3.7 Utterances ending in minor boundaries and Utts continuing after major boundaries

Utterance (Utt) has been defined according to its prosodic structure, namely, as a discourse or information unit ending in a major boundary. There are, however, some exceptions to it: On the one hand, there are Utts ending in what seems to be perceived as minor boundaries; on the other hand, there are cases where a major IM does not necessarily indicate that a preceding series of IMs has come to its end, thus forming a coherent Utt. These two options will be discussed in the following two sections (Section 3.7.1 and Section 3.7.2 respectively).

3.7.1 *Utterances ending in minor boundaries*

As mentioned, there are cases where what is perceived as a minor boundary actually ends an Utt. Among these Utts, we can distinguish a few main categories that a study of the present corpus has revealed: isotonic Utts (Section 3.7.1.1); backchannels (Section 3.7.1.2); greetings and courtesy phrases (Section 3.7.1.3); suspended Utts (Section 3.7.1.4). It should be noted that not all Utts in the categories listed below (save suspended ones) end in a minor boundary. Their ending in a minor boundary is thus optional.

3.7.1.1 *Isotony*

Isotony, or intonational (tonal) parallelism, is the repetition of (part of) the previous prosodic contour in a following PM (Izre'el & Mettouchi, 2015, pp. 23–25; following Du Bois, 2006; Wichmann, 2000, Section 4.3). When a final PM in a PS (Utt) copies the boundary tone of the previous PM, an Utt may end in a tone similar to that of a minor boundary. Isotony may sometimes occur in lists and their like. Example (24), an excerpt taken from instructions given to a group of soldiers by their commander during a briefing before starting their night watch, illustrates this. The instructions are part of a series of actions when a suspicious person approaches.

(24) aˈʦor ve hizdaˈhe |
stop and identify.youself
‘“Stop and identify yourself!”’
aˈʦor o ʃe aˈni joˈre |
stop or that I shoot
‘“Stop or I shoot!”’
ʃikˈʃuk be=ʃiˈʃim maaˈlot |
rattling in=sixty degrees
‘Rattling in 60º,’
ˈjeri bejn=kavaˈnot |
shooting between=sights
‘shooting between sights;’ [source: P423_1_sp5_002–005]

The commander pauses after the last IM, enabling the soldiers to ask questions or respond. Soldiers indeed ask questions, the commander answers them and then he resumes his briefing by saying *aχar kaχ* 'later', 'then' (source: P423_1_sp5_010). Still, the final boundary tone of the cited Utt is perceived as minor, being a copy of the previous PMs.

Minor boundary at the end of lists may signal that the list is incomplete. A similar boundary tone may further signal incompleteness also in other cases, including an Utt consisting of only one IM. In (25), the speaker is describing the benefits of staying in a new hotel:

(25) aru'χot ka'ful |
meals double
'Double meals…' [source: OCD_2_sp1_034]

She mentions only one benefit, suggesting by the minor boundary that there are others like this. In such cases, the non-falling tone, which basically carries the idea of continuity (Section 2.1.1), seems to indicate the notion of "et cetera", "or the like". Supporting evidence for this meaning carried by prosody is supported by SMs encapsulated by isotonic PSs:

(26) am'ra=li ba'rur ʃe at jeχo'la lih'jot kan | ve ze | ve ze |
she.said=to.me clear that you.SGF can be here and this and this
'She said to me: "of course you can stay here", and so on and so forth…'
[source: Y34_sp1_171–173]

The final two IMs in (26) consist of conjunctions and pro-forms indicating the notion of "and other things", "and so on". Still, there are cases of isotony that seem to result merely by attraction of form, as is the case with Example (27), where the speaker is telling about lice. A pause of 0.7" s following this Utt, before her interlocutor responds and changes subject, indicates that there was no intention to continue it.

(27) oto'matit hit'χalti l=hitga'red bete'ruf | hi'gati ha'bajta | v
automatically I.began to=scratch in.craziness I.reached homeward and
hista'rakti | ki ze=ha'ja kolkaχ do'χe |
I.combed because it=he.was so repelling
'Automatically I started scratching frantically, I reached home, and I combed (my hair), because it was so repelling…' [source: C711_0_sp1_202–205]

3.7.1.2 *Backchannels*

Backchannels – either paralinguistic (Example (28)) or linguistic (Example (29)) – can be pronounced by "continuing" boundary tones, suggesting that the main interlocutor is to continue speaking disregarding the intervention of the listener.

(28) Sp1: bekaˈrakurum | ˈjeʃ et=ha= | eːmp | et=ha= | tiˈra ʃel=
in.Karakorum EXT ACC=DEF= uhm ACC=DEF= castle of=
e: | (0.388) ˈʤiŋgis ||
uh Genghis
'In Karakorum there is the castle of Genghis.'
Sp2: mˈm̥m |
'Mhm.' [source: OCh_sp1_116–121]

(29) Sp1: beraˈmot ʃoˈnot mtugmaˈlim aˈχeret || (1.010) bimekoˈmot
in.levels different rewarded otherwise in.places
ʃoˈnim |
different
'In different levels (people) are rewarded differently. (i.e.,) In different places.'
Sp2: okej |
'Okay.' [source: P931_2_sp1_072–073; sp2_03]

3.7.1.3 *Greetings and courtesy phrases*

At times, greetings and courtesy phrases may carry boundary tones that sound similar to a minor boundary. In (30), the speaker is about to get off a car, thanking and greeting his friends, who brought him back home.

(30) toˈda banˈot | (0.221) nsiˈa toˈva |
thanks girls trip good
'Thank you, girls. (Have a) good trip.' [source: OCD_3_sp1_077–078]

3.7.1.4 *Suspended utterances*

The most widespread category among Utts ending in minor boundaries is the category of suspended Utts. Suspended IMs are syntactically and semantically incomplete units. The minor boundary at their end suggests to the interlocutor that the speaker has not finished this Utt and seems to be intending to add another IM in succession. This intention is not immediately satisfied, although it can be satisfied later. Thus, recognition of suspension is made by segmental features rather than by prosodic ones (e.g., abrupt final syllable, glottal stop ending), which are signals for truncation (= fragmentary unit; Section 2.1.1, end).

The speaker in (31) has started to speak, yet notes that he needs to supply some background information for his interlocutor.

(31) ba'jom ʃe'bo |
in.the.day in.which
‘In the day during which …’
<inhale and sigh> (2.474)
a'naχnu 'garnu be=e'lat ||
we we.lived in=Eilat
‘We lived in Eilat.’ [source: P931_2_sp1_046–048]

He continues with the background story, and never gets back to the suspended Utt. In contrast, the speaker in (32) likewise leaves her Utt unfinished because her interlocutor, Sp1, has by responding overlapped her question:

(32) Sp2: ejn baa'ja [ʃel e |]
NEG.EXT problem of uh
‘There is no problem that…’
Sp1: [lo ||]
NEG
‘No.’
lo baa'ja ||]
NEG problem
‘No problem.’ [source: C711_0_sp2_068; sp1_067–068

In neither (31) nor (32), there are prosodic indications for either truncation or neglect of the current Utt (or IM), since the respective PM ends in a minor (continuing) boundary tone. Therefore, the speakers might have well continued their suspended Utt if they felt the need to complete it. The choice of the term *suspension* for this type of Utts reflects this situation.

While many of the suspended IMs end the Utt they are part of (either as singletons or as ending series of minor IMs), there are cases where suspended IMs will not be regarded as (ending) suspended Utts. This is the case where suspension is made for (unplanned) repetition or syntactic or semantic repair, enabling the speaker to end his initiated Utt and convey the information in whole. We have seen one case of such unplanned repetition (within a single IM) in (16) above. Another illustration for suspension within an Utt is Example (33), where the speaker restarts her Utt for repair:

(33) ve hem gam | at o'meret ʃe hem do'mim ||
and they also | you.SGF say that they resembling
‘And they are also… you say that they look alike.’
[source: Y311_sp1_190–191]

The boundaries ending a suspended IM are genuine minor boundaries, in that the speaker itself produces them as such in his intention to continue speaking. On the other hand, the previous three categories (Sections 3.7.1.1–3) are only apparent minor boundaries, as – at least in some of the cases described – the speaker does not mean to continue the Utt ending this way. Further research is needed to establish prosodic differences between these types of contours.

3.7.2 *Major boundaries in mid-utterance position*

In contrast to cases where an Utt ends in a minor boundary (or what sounds like one) (Section 3.7.1), there are cases where a major boundary does not indicate the ending of a preceding (series of) minor IM(s). This is the case with inserts (Section 3.7.2.1) and split Utts (Section 3.7.2.2).

3.7.2.1 *Inserts*

The most obvious inserts are discourse markers, which come in the middle of an Utt and can end either with a minor boundary or with a major boundary, whether signaled by either a falling tone, as in (34) and (36), or a high rise ("appeal"; Section 2.1.1) as in (35).

(34)

im=af	gaˈdol |	ve |	im=	**ata-joˈdea** ||	ˈsal	kaze	leˈmala |
with=nose	big	and	with=	**you=know**	Basket	like.this	above

laˈsim	et=ˈkol=ha= |	peklaˈot |
to.put	ACC=all=DEF=	luggage

'With a big nose, and with you know sort of a basket above, to put all the luggage.' [source: OCh_sp1_199–204]

In (34), the discourse marker *ata jodea* 'you know' is inserted immediately following the first word in the second IM of an Utt, a proclitic preposition. The prepositional phrase continues immediately after the major boundary of the inserted phrase. In (35), the speaker is describing the home of some friends, which both he and his interlocutor know.

(35)

eχ	ʃe	niχnaˈsim |	ken / (0.847)	jeʃ	misdaˈron |
how	that	entering	yes	EXT	corridor

'As you enter – yes? – there is a corridor.' [source: C842_sp1_142–144]

The insert *ken* / 'yes?' comes between the fronted locative phrase and the existential clause. While in (34) and (35) the inserts are discourse markers, Example (36) illustrates a substantive insert:

(36) na'gid aχo'ti | kʃe hi=jo'tset im=ana'ʃim | **im=baχu'rim ||** (0.663)
Say my.sister when she=go.out with=people **with=guys**
hi=jot- hi=jo'tset le=mata'ra msu'jemet ||
she=ERR she=go.out to=goal specific
'Say, when my sister dates people – (that is,) with guys – she dates (them) for a specific reason.' [source: P423_2_sp1_ 041–044]

In (36), the insert serves to correct or make clearer the speaker's use of *anaʃim*, which is the general term for 'people' and may be misinterpreted, whereas *baχurim* is usually used for 'younger men' and fits more the context of dating.

3.7.2.2 *Split utterances*

As mentioned above (Section 3.7.1.4), suspended Utt may be resumed later in the conversation. Usually in such cases, when a suspended topic of conversation is resumed, the speaker starts a new Utt. In rare cases, however, there are acoustic data that support the analysis of split Utts, as is the case in (37).

(37) Sp1: aχ'ʃav || e | a'ni jeχo'la leha'gid=laχ | ʃe ani=maki'ra | zug
Now uh I can tell=you that I=recognize couple
e | na'suj |
uh married
'Now, uh I can tell you that I know uh a married couple,'
he || ja'fe || titχad'ʃi ||
oh beautiful enjoy.the.new
'Oh! Nice! Enjoy!'
Sp2: ze ha=ba'nim ka'nu=li ||
this DEF=sons bought=to.me
'The boys bought it for me.'
Sp1: ja'fe ||
beautiful
'Nice!'
Sp2: eze χamu'dim ||
which cute.PL
'They are cute.'
Sp1: na'χon ||
right
'Right.'
mak'sim ||
charming
'Charming.'

Sp1: ʃe em | ʃe ʃnejˈhem haˈju nesuˈim | az hem
that uhm that both.of.them they.were married so they
baalej=oto=inteˈres liʃmor=al=ze be=sodiˈjut | ve | hi |
owners.of=same=interest to.guard=on=this in=secrecy and she
hitpanˈta | ve nigˈmar ha=ˈkeʃer ||
she.became.free and it.be.finished DEF=bond
'that uhm that both of them were married, so they had the same interest to keep it a secret, and (then) she became single and the relationship ended.' [source: Y311_sp1_030–048; sp2_004–005]

The insert, marked here by smaller font and indents, includes a complete conversation of eight Utts. The insert is initiated by Sp1, who is the one telling her friend about the married couple she knows. She stops the main message almost in its beginning, where the minor boundary indicates that it is to be continued. For changing subject, she softens her voice loudness considerably. Her interlocutor responded with the same low volume, and this quiet secondary conversation lasts until its end. The main speaker now resumes her first topic. The voice volume rises again. Moreover, the suspended Utt resumes as it would be uttered in its original location, without repeating the subject matter and by using the particle *ʃe* 'that', opening a relative clause.

4. Conclusion

An *Utterance* (Utt) is a discourse unit that can be defined as an *information set* (IS), consisting of one or more *information modules* (IMs). As such, it consists of one or more *segmental modules* (SMs) encapsulated by their respective *prosodic modules* (PMs).

An Utt communicates a single, integrated information package. An Utt is the domain of the clause. For substantive units, an Utt consists minimally of a clause. In addition to a single clause, an Utt can include clausal or non-clausal regulatory components, inserts, repeats, intra-thematic repairs, including suspended or truncated IMs. Furthermore, an Utt can include a clause cluster. As such, it may be regarded as the domain of a spoken sentence, which is defined as the syntactic unit consisting minimally of a clause (Section 3.5.2). Finally, an Utt can consist of non-clausal material, mostly regulatory, including paralinguistic elements (cf. Section 3.4; Section 3.7.1.2, Example (28)).

It has been suggested, that the Utt is the best candidate to be regarded as the basic unit of spoken language. The arguments for preferring it over the IM can be summarized as follows:

a. *Phonetic*: A unit ending in a falling boundary tone will be regarded as primitive and hence primary (Section 2.1.1).
b. *Phonological (prosodic)*: A major boundary indicates finality, whereas a minor boundary indicates same-speaker continuity (Section 1.3.1.1).
c. *Informational*: An Utt conveys an integrated informational unity (Section 2).
d. *Syntactic*: The Utt is the domain of the basic unit of syntax, namely, clause (Sections 3.1–3.3).
e. *Pragmatic*: The Utt is a speech-act unit and has an internal structure in accordance with the Language into Act Theory (L-AcT) (Cresti, 2000; and subsequent publications; Raso & Moneglia, 2014).
f. *Conversational*: The boundary of an Utt signals prosodic completion, which indicates a transition relevance place (TRP) (Section 2.1.2).

Acknowledgements

Heliana Mello and Amina Mettouchi read a previous version of this chapter. I thank both of them for useful comments.

References

Amir, N., Silber-Varod, V., & Izre'el, S. (2004). Characteristics of intonation unit boundaries in spontaneous spoken Hebrew: Perception and acoustic correlates. In B. Bel, & I. Marlien (Eds.), *Speech prosody 2004* (pp. 677–680). Nara, Japan: ISCA.

Bolinger, D. (1972 [1964]). Around the edge of language: Intonation. In D. Bolinger (Ed.), *Intonation: Selected readings* (pp. 19–29). Harmondsworth: Penguin Books.

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

Bossaglia, G., Mello, H., & Raso, T. (this volume). Illocution as a unit of reference for spontaneous speech: An account for insubordinated adverbial clauses in Brazilian Portuguese. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Brazil, D. (1995). *A grammar of speech*. Oxford: Oxford University Press.

Chafe, W. (1993). Prosodic and functional units of language. In J. Edwards & M. D. Lampert (Eds.), *Talking data: Transcription and coding in discourse research* (pp. 33–43). Hillsdale, NJ: Lawrence Erlbaum Associates.

Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago, IL: The University of Chicago Press.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.

CoSIH. (2012). *The corpus of spoken Israeli Hebrew (CoSIH)*. Retrieved from <cosih.com/english/index.html>

Cresti, E. (2000). *Corpus di Italiano parlato*. Florence: Accademia della Crusca.

Croft, W. (1995). Intonation units and grammatical structure. *Linguistics*, 33, 839–882. https://doi.org/10.1515/ling.1995.33.5.839

Croft, W. (2007). Intonation units and grammatical structure in Wardaman and English. *Australian Journal of Linguistics*, 27, 1–39. https://doi.org/10.1080/07268600601172934

Cruttenden, A. (1997). *Intonation*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139166973

Crystal, D. (2008). *A dictionary of linguistics and phonetics* (6th ed.). Oxford: Blackwell. https://doi.org/10.1002/9781444302776

Debaisieux, J-M. (Ed.). (2013). *Analyses linguistiques sur corpus: Subordination et insubordination en français*. Cachan: Hermès & Lavoisier.

Debaisieux, J-M., & Martin, P. (this volume). Syntactic and prosodic segmentation in spoken French. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Den, Y., Koiso, H., Takehiko, M., Maekawa, K., Takanashi, K., Enomoto, M., & Yoshida, N. (2010). Two-level annotation of utterance-unit in Japanese dialogs: An empirically emerged scheme. *Proceedings of LREC 2010*, 2103–2110.

Du Bois, J. W. (2006). Representing discourse. Retrieved from <http://www.linguistics.ucsb.edu/projects/transcription/representing>

Du Bois, J. W., Cumming, S., Schuetze-Coburn, S., & Paolino, D. (1992). *Discourse transcription*. Santa Barbara, CA: Department of Linguistics, University of California.

Du Bois, J. W., Cumming, S., Schuetze-Coburn, S., & Paolino, D. (1993). Outline of discourse transcription. In J. A. Edwards & M. D. Lampert (Eds.), *Talking data: Transcription and coding in discourse research* (pp. 45–89). Hillsdale, NJ: Lawrence Erlbaum Associates.

Evans, N. (2007). Insubordination and its uses. In I. Nikolaeva (Ed.), *Finiteness: Theoretical and empirical foundations* (pp. 366–431). New York, NY: Oxford University Press.

Féry, C. (2017). *Intonation and prosodic structure*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781139022064

Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 134–184). Cambrige: Cambridge University Press. https://doi.org/10.1017/CBO9780511620874.003

Halliday, M. A. K. (2014). *Halliday's Introduction to Functional Grammar*. Fourth edition revised by Christian M. I. M. Matthiessen. London and New York: Routledge. https://doi.org/10.4324/9780203783771

Harris, Z. S. (1951). *Structural linguistics*. Chicago, IL: The University of Chicago Press.

Hirst, D., & Di Cristo, A. (Eds.). (1998). *Intonation systems: A survey of twenty languages*. Cambridge: Cambridge University Press.

Hockett, C. F. (1958). *A course in modern linguistics*. New York, NY: Macmillan. https://doi.org/10.1111/j.1467-1770.1958.tb00870.x

Inbar, A. (2016). Is subordination viable? The case of Hebrew ʃɛ 'that'. *CHIMERA: Romance Corpora And Linguistic Studies*, 3(2), 287–310.

Iwasaki, S. (1993). The structure of intonation units in Japanese. In S. Choi (Ed.), *Japanese/Korean linguistics* (Vol. 3, pp. 39–53). Stanford CA: CSLI.

Iwasaki, S., & Tao, H. (1993). *A comparative study of the structure of the intonation unit in English, Japanese, and Mandarin Chinese*. Paper presented at the Annual Meeting of the Linguistics Society of America, Los Angeles, CA. Retrieved from <https://www.researchgate.net/publication/241563041_A_Comparative_Study_of_the_Structure_of_the_Intonation_Unit_in_English_Japanese_and_Mandarin_Chinese>

Izre'el, S. (2002). The corpus of spoken Israeli Hebrew: Textual samples. *Leshonénu*, 64, 289–314.

Izre'el, S. (2005). Intonation units and the structure of spontaneous spoken language: A view from Hebrew. In C. Auran, R. Bertrand, C. Chanet, A. Colas, A. Di Cristo, C. Portes, A. Reynier, & M. Vion (Eds.), *Proceedings of the IDP05 international symposium on discourse-prosody interfaces*. Retrieved from <https://www.academia.edu/229811/Intonation_Units_and_the_Structure_of_Spontaneous_Spoken_Language_A_View_from_Hebrew>

Izre'el, S. (2012). Basic sentence structures: A view from spoken Israeli Hebrew. In S. Caddéo, M.-N. Roubaud, M. Rouquier, & F. Sabio (Eds.), *Penser les langues avec Claire Blanche-Benveniste* (pp. 215–227). Aix-en-Provence: Presses Universitaires de Provence.

Izre'el, S. (2018a). Unipartite clauses: A view from spoken Israeli Hebrew. In M. Tosco (Ed.), *Afroasiatic: Data and perspectives* (pp. 235–259). Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.339.13izr

Izre'el, S. (2018b). Syntax, prosody, discourse and information structure: The case for unipartite clauses. A view from spoken Israeli Hebrew. *Revista de Estudos da Linguagem*, 26(4), 1675–1726. Retrieved from <http://periodicos.letras.ufmg.br/index.php/relin/article/view/13036/pdf>

Izre'el, S., & Mettouchi, A. (2015). Representation of speech in CorpAfroAs: Transcriptional strategies and prosodic units. In A. Mettouchi, M. Vanhove, & D. Caubet (Eds.), *Corpus-based studies of lesser-described languages: The CorpAfroAs corpus of spoken AfroAsiatic languages* (pp. 13–41). Amsterdam: John Benjamins.

Kibrik, A. A., Korotaev, N. A., & Podlesskaya, V. I. (this volume). Russian spoken discourse: Local structure and prosody. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Kibrik, A. A., & Podlesskaya, V. I. (2006). Problema segmentacii ustnogo diskursa i kognitivnaja sistema govorjashchego [Segmentation of spoken discourse and the speaker's cognitive system]. In V. D. Solovyev, (Ed.), *Kognitivnye issledovanija* (Vol. 1, pp. 138–158). Moscow: Institut psixologii RAN.

Kibrik, A. A., & Podlesskaya, V. I. (2008). Is sentence viable? In *The third international conference on cognitive science. June 20–25, 2008, Moscow, Russia. Abstracts* (Vol. 1, pp. 84–85). Moscow: IP RAN.

Kibrik, A. A., & Podlesskaya, V. I. (Eds.). (2009). *Rasskazy o snovidenijax: korpusnoe issledovanie ustnogo russkogo diskursa* [Night dream stories: A corpus study of spoken Russian discourse]. Moskva: Iazyki Slavjanskyx Kultur.

Liddicoat, A. J. (2007). *Introduction to conversation analysis*. London: Continuum.

Lieberman, P. (1967). *Intonation, perception and language*. Cambridge, MA: The MIT Press.

Lieberman, P., & Blumstein, S. E. (1988). *Speech physiology, speech perception, and acoustic*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139165952

Lyons, J. (1968). *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139165570

Martin, P. (2015). *The structure of spoken language: Intonation in romance*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139566391

Matsumoto, K. (2003). *Intonation units in Japanese conversation: Syntactic, informational and functional structures*. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.65

Mithun, M. (2008). The extension of dependency beyond the sentence. *Language*, 83, 69–119. https://doi.org/10.1353/lan.2008.0054

Raso, T., & Moneglia, M. (2014). Notes on Language into Act Theory (L-AcT). In T. Raso & H. Mello (Eds.), *Spoken corpora and linguistic studies* (pp. 468–495). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61

Sacks, H., Schegloff, E. A., & Jefferson, G. L. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 696–735. https://doi.org/10.1353/lan.1974.0010

Selkirk, E. (2001). The syntax-phonology interface. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (2nd ed., Vol. 23, pp. 895–899). Amsterdam: Elsevier. https://doi.org/10.1016/B0-08-043076-7/02958-2

Shor, L. (2016). Cognitive and interactional motivations for prosodic phrasing: A corpus-based analysis of the clause in spoken Israeli Hebrew. *CHIMERA: Romance Corpora and Linguistic Studies*, 3(2), 325–343.

Silber-Varod, V. (2011). The SpeeCHain perspective: Prosodic-syntactic interface in spontaneous spoken Hebrew (Unpublished doctoral dissertation), Tel-Aviv University, Israel. Retrieved from <http://www.openu.ac.il/Personal_sites/vered-silber-varod/download/Vered%20Silber-Varod%20Dissertation-7.pdf>

Sornicola, R. (2006). Topic and comment. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd ed., pp. 766–773). Oxford: Elsevier. https://doi.org/10.1016/B0-08-044854-2/00594-0

Szczepek Reed, B. (2011). *Analysing conversation: An introduction to prosody*. Houndmills: Palgrave Macmillan. https://doi.org/10.1007/978-1-137-04514-0

Tao, H. (1996). *Units in Mandarin conversation: Prosody, discourse, and grammar*. Amsterdam: John Benjamins. https://doi.org/10.1075/sidag.5

Warren, P. (2016). *Uptalk: The phenomenon of rising intonation*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316403570

Wichmann, A. (2000). *Intonation in text and discourse: Beginnings, middles and ends*. Harlow: Pearson Education.

Wouk, F. (2008). The syntax of intonation units in Sasak. *Studies in Language*, 32, 137–162. https://doi.org/10.1075/sl.32.1.06wou

CHAPTER 3

Prosody and the organization of information in Central Pomo, a California indigenous language

Marianne Mithun
University of California, Santa Barbara

In some theoretical frameworks, it is assumed that prosodic structure is a direct reflection of syntactic structure. Close examination of unscripted speech confirms that though the two often work in concert, they are distinct. Prosodic structure differs from grammatical structure in some fundamental ways. Prosody (pitch, intensity, rhythm) involves continua and can be more responsive to certain subtle differences in cognitive state, discourse context, and interactive goals. Grammar (morphology and syntax) can mark more distinctions, but these are categorical and conventionalized: an affix is either present or absent; one constituent either precedes or follows another. Here some prosodic structures, their functions, and their relation to grammatical structures are discussed with examples from Central Pomo, a language indigenous to Northern California.

Keywords: intonation unit, prosodic sentence, Central Pomo, topicalization, clause combining

1. Introduction

It has sometimes been assumed that prosodic structure, involving such distinctions as pitch, intensity, and rhythm, is a direct reflection of syntactic structure. Closer examination of unscripted speech shows that though the two often work in concert, neither can be predicted directly from the other. Each adds meaning of its own. Relations between the two are illustrated here with speech in Central Pomo, a language indigenous to Northern California. The material is drawn from unscripted conversation and narrative within conversation recorded over a period of nine years from residents of the three Central Pomo communities: Frances Jack and Alice Elliott of the Hopland Rancheria, Florence Paoli and Salome Alcantra of the Yokayo Rancheria, and Eileen Oropeza, Winifred Leal, and Jesse Frank of the Point Arena/Manchester Rancheria. In Section 2 the notion of Intonation

https://doi.org/10.1075/scl.94.03mit
Additional files available at http://doi.org/10.1075/scl.94.03mit/audio

Unit is introduced; in Section 3 prosodic and syntactic sentences are compared; in Section 4 the prosody of nominal phrases is described, and in Section 5 the prosody of clause-linking constructions is discussed. As will be seen, some of the differences are due to fact that while the prosodic features are continua, the grammatical structures are inherently categorical.

2. Intonation Units and the packaging of information

The basic unit of prosodic analysis used here is the Intonation Unit (prosodic phrase), characterized by an initial pitch reset, a continuous pitch contour, and an identifiable terminal contour, often but not necessarily bordered with pauses. The free translation of one brief account, which was originally in Central Pomo, is presented in (1). Each line represents a separate Intonation Unit. Punctuation reflects prosodic structure, with commas for non-final pitch contours, periods for final terminal contours, and – for truncation. (Truncation is generally accompanied by no fall at all in pitch.)

(1) My Brother's Wife
1. A long time ago,
2. my older brother married,
3. married a woman by the name of Maggie.
4. Um at that time,
5. um–
6. on the west side of the lake,
7. on the mountainside,
8. he was cutting his wood,
9. chopping wood.
10. The woman was there too,
11. when she got her period.
12. When Indian people had their period,
13. they weren't supposed to look at that water.
14. They say a thing,
15. a monstrous thing,
16. a wild thing,
17. like a snake,
18. in there they say,
19. a (menstruating) woman would see.
20. That woman saw that they say.
21. They say she,
22. sort of,
23. got sick.

24. From then on they say,
25. she began a song.
26. Sang a song.
27. My brother would sing that song too.
28. That woman,
29. died early.

The Intonation Units are often but not always bounded by pauses. The lengths of pauses, of Intonation Units, and of prosodic sentences in a section of (1) are given in seconds in (2).

(2) Central Pomo Intonation Units: Frances Jack, speaker p.c.

IU#	*Pause*	*Free translation*	*IU length*	*S length*
12	**0.636**	When Indians had their period,	1.903	
13	0.469	they weren’t supposed to look at that water.	3.088	5.460
14	**0.629**	They say a thing,	0.917	
15	0.194	a monstrous thing,	1.524	
16	0.448	a wild thing,	0.918	
17	0.138	like a snake,	1.375	
18	0.475	in there they say,	0.736	
19	0.353	a woman would see.	1.639	8.757

The primary defining feature of Intonation Units is pitch: an initial pitch reset, a continuous pitch contour, and an identifiable terminal contour. A typical Intonation Unit from (1) is shown in (3).

(3) *Mú:l dó: mu:l má:t̯a ’el maqó-w.*
that HEARSAY that woman the see-PFV
‘The woman saw that they say.’

Intonation Units are also often characterized by a continuous intensity contour. In Figure 1, the lower line shows the pitch trace of the sentence in (3), and the upper line the intensity. The numbers at the bottom of the figure represent time.

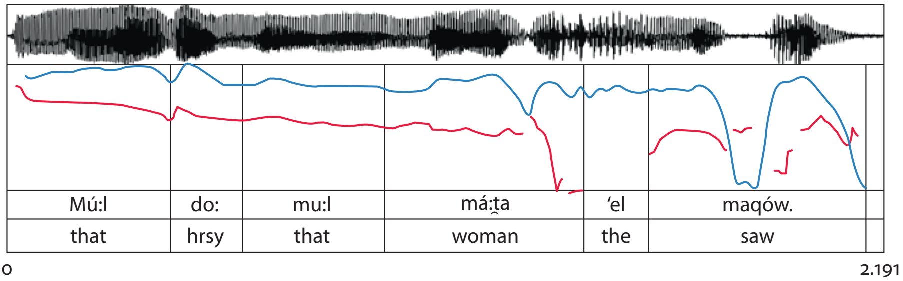


Figure 1. Pitch and intensity contours

There is also typically an overall declination in pitch, visible here as well in the downward slope of the red pitch trace. There is generally less anacrusis or final lengthening in Intonation Units among these speakers.

Intonation Units can occur on their own or in combination to form prosodic sentences, recognizable by an initial full pitch reset, usually a continuous declination in pitch over the whole (with possible smaller pitch resets at the beginning of component Intonation Units), and a final terminal contour. They are often, though not necessarily, preceded by a pause of 600–900 milliseconds, as can be seen in (2) in bold. The opening syntactic sentence in (1) was matched by a prosodic sentence, which consisted of three smaller Intonation Units. Again the punctuation reflects prosody.

(4) *Mu:l ʼma šé:mi b-bal,*
that FACT long ago this
'A long time ago,'
kí:ki báʼdu-w,
my brother marry-PFV
'my brother married,'
Maggie ši báʼdú-čʼ ʼe.
Maggie name marry-SML COP
'married a woman named Maggie.'

The pitch and intensity traces for this sentence can be seen in Figure 2. Each Intonation Unit shows a declination in pitch, followed by a pause and a pitch reset at the beginning of the next Intonation Unit. The pause before the second Intonation Unit was 0.2751 seconds, and that before the third was 0.7169 seconds. The sentence as a whole shows an overall declination in pitch and a final terminal pitch contour.

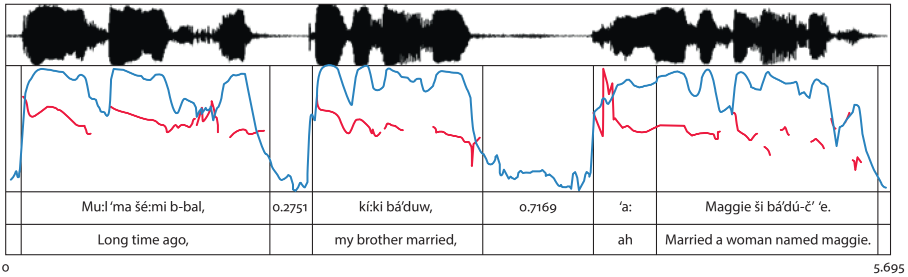


Figure 2. Prosodic sentence

The prosodic structuring of spontaneous speech is not random. As described in detail by Chafe (1979, 1984, 1987, 1988, 1992, 1994), Intonation Units generally correspond to cognitive entities: Speakers tend to introduce no more than one significant new piece of information at a time. This may be a participant, an event, a time, a place, an elaboration, etc. Such structuring is apparent in (4) above, where each line represents an Intonation Unit and also a significant new idea.

Given or accessible information, that is, information already active in the mind of the speaker or semi-active from earlier mention or association with active or semi-active information, may be combined with new information in an Intonation Unit. An example can be seen in the third Intonation Unit of (4): ‘He married a woman named Maggie’. The idea of marrying was given, just introduced in the previous Intonation Unit, so the one new idea was Maggie. Chafe describes the Intonation Unit as reflecting a single focus of consciousness. Larger pauses reflect extra thought, as visible in Figure 2 before the hesitation *a:* and before the second line in (5).

(5) IU Pause
21 0.808 They say she,
22 **1.675** sort of,
23 0.205 got sick.

Also as described by Chafe, the structuring of spontaneous speech into prosodic sentences is not random. Prosodic sentences tend to represent semantic entities, generally expressing one event or state. They can vary in length, here between 2.229 seconds and 8.757 seconds. Longer pauses can be seen at major discourse breaks, as in (6).

(6) IU Pause
27 0.882 My brother used to sing that song too.
28 **1.139** That woman,
29 0.059 uh died early.

When we turn to conversation, the same general patterns can be seen. Intonation Units still show one new idea at a time. In (7), as some speakers were eating cake, one commented that it tasted good, then looked at her friend Frances and added that it tasted good to Frances.

(7) Central Pomo Conversation
Pause
FP *Bal yawál q’dí qa:-t̯’á:-d=a,*
this all good biting-sense-IPFV.SG-IMM
‘This is tasting good to everyone,’
FP, SA 0.922 (laughter)
FP *Frances q’dí qa:-t̯’á:-d=a,*
Frances good biting-sense-IPFV.SG=IMM
‘It tastes good to Frances,’

But conversation also raises additional issues. Do we count pauses between *all* speech events, or only between the Intonation Units of a single speaker? Do we count laughter as pauses? Where do backchannel responses like *mhm* fit? As is well known, pause length between turns is both highly cultural and individual. And of course in everyday life, the distinction between monologue and conversation is not always clear cut. Monologue of varying lengths is most often embedded in conversation.

In the examples seen so far, prosodic sentences correspond to syntactic sentences. Both express one event or state. Intonation Units often match syntactic constituents, but the level of the constituents varies: a temporal phrase ('a long time ago'), a locative phrase ('on the mountainside'), a noun phrase ('a monstrous thing'), a verb ('got sick'), a verb phrase ('chopping wood'), a clause ('my older brother married'), etc. In fact the division into prosodic units is not fully accounted for by syntactic structure. It more directly reflects the status of the information conveyed in the mind of the speaker and hearer.

3. Sentence boundaries

In many models of syntax, divisions between sentences are categorical. But the dimensions of prosody are continua: rhythm, pitch, and intensity. And in spoken language, the strength of prosodic divisions between sentences can be a matter of degree. Example (8) consists of two syntactic sentences.

(8)	*ʼa:*	*ṭo*	*bédah–*	*béda=h̯ṭow*	*bé:=yo-w*	*dá:ʼdu-w*	*čʰó-w.*
	1SG.AGT	CONTR	here–	here=from	away=go-PFV	want-PFV	be.not-PFV

'I don't want to go away from here.'

Béda	*ʼa:*	*q'lá:-w=ʼkʰe.*
here	1SG.AGT	die-PFV=FUT

'I will die here.'

It is worth noting that the first sentence is actually syntactically complex, with matrix clause 'I don't want' and complement clause 'to go away from here'. It was still packaged as a single prosodic sentence, with regular declination in pitch over the whole, ending in a final terminal fall.

There is more to this sequence, however. The two sentences in (8) are related semantically, though that relation is not signaled by syntactic structure. It is, however, indicated by prosody. As just noted, the first sentence ended with a terminal fall in pitch and was followed by a pause, as would be expected of a prosodic sentence. But the second did not begin with a full pitch reset (see Figure 3).

The apparent pitch peak in the second Intonation Unit is just the glottal stop of *q'lá:w'kʰe*.

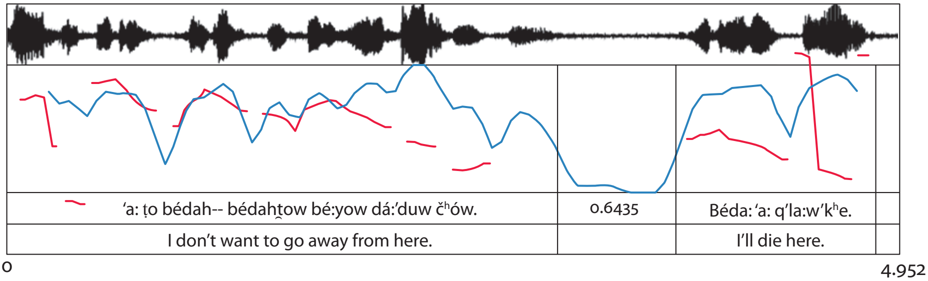


Figure 3. Related sentences

Friends sitting around in the reception area of an office were wondering where they might find something to drink. One volunteered the remark in (9).

(9) *Qʰá t̯ika mene mu:l meṭ'*–
water you know that such
'Water you know,'
uh–
čó:-č'i-w-a'-ya-m=ma
keep-IPFV.PL-PFV-IPFV.PL-PASS-MULT.AGT=FACT
'they keep'
ni:n čók ṭʰíy-a:y=li.
so jug big-DISTR=with
'in big jugs like this.'
Meṭ' k'íw s-t̯'á-man.
such cold sucking-sense-PARTICULAR
'It tastes cold.'

The first sentence 'They keep water in big jugs like this' ended with a final terminal fall in pitch. It was followed by a pause of 0.3730 seconds before the second sentence 'It tastes cold', which began with a full pitch reset. The two were packaged as clearly distinct sentences both prosodically and syntactically (see Figure 4).

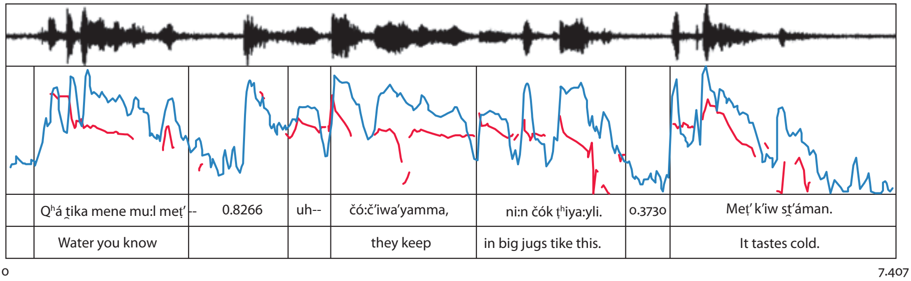


Figure 4. Distinct sentences: Terminal fall, pause, pitch reset

But prosodic sentences do not always match syntactic sentences. After (9), another person asked ‘Here?’. The first speaker responded ‘They keep it here in the office for the workers’. She then added the comment in (10).

(10) *Čók ṭʰíyay čʰmáh-duwa:dan meṭ’,*
jug large-DISTR sit.on-IPFV such
‘There are large jugs sitting up there,’
k’íw s-t̯’a-w.
cold sucking-sense-PFV
‘it tastes cold.’

Here the second sentence began with a pitch reset, but it was not preceded by a pause (see Figure 5). The fact that the water was cold was not new. The difference in information status was conveyed prosodically but not syntactically.

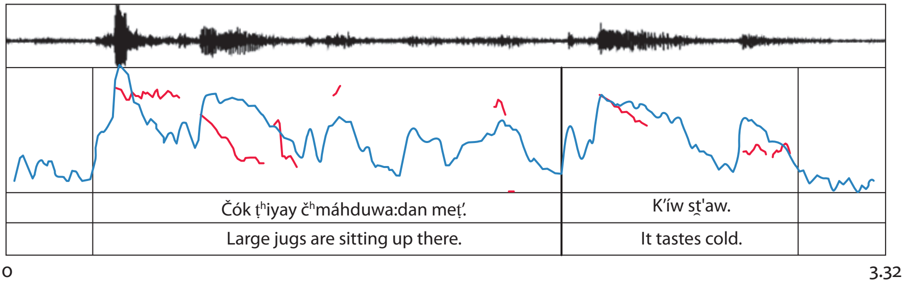


Figure 5. Terminal fall and pitch reset without pause

Central Pomo contains some special rhetorical structures that relate sequences. A common one is a kind of couplet construction, in which a statement is made, then is followed by a restatement with some twist: a change in word order, vocabulary, voice, etc. These constructions show a distinctive echoing intonation, with repetition of the rhythm and pitch pattern. An example from the story in (1) is in (11).

(11) *Mída=ht̯ow do: mu:l,*
there=from HRS that
‘From then on they say she’
kʰé be:-yú’-č’i-w;
song orally-begin-SML-PFV
‘began a song;’
kʰé če:nó-w.
song sing-PFV
‘sang a song.’

The Intonation Units ‘began a song’ and ‘sang a song’ were spoken with the same timing and pitch contour (see Figure 6).

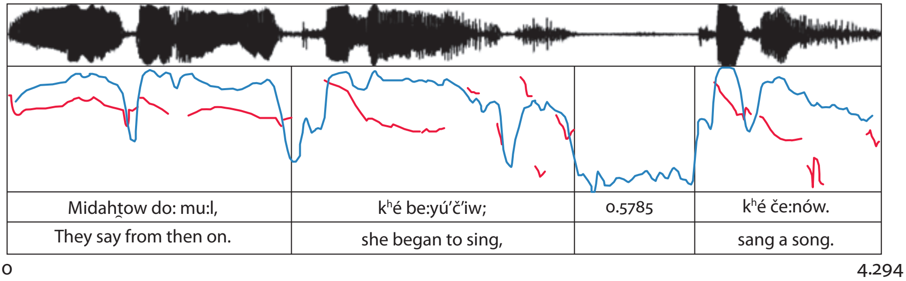


Figure 6. Couplet structure

A similar construction occurred earlier in (1), repeated here as (12).

(12) *t̯í:kʰe- háy yhé:-n,*
own wood do-IPFV.SG
'he was cutting his wood,'
háy pʰqʰám.
wood swinging-chop
'chopping wood.'

A related construction consists of two parallel sentences in which the second elaborates on the first. Example (7), repeated here as (13), is such a construction.

(13) *Bal yawál q'dí qa:-t̯'á:-d=a,*
this all good biting-sense-IPFV.SG=IMM
'This is tasting good to everyone;' (laughter)
Frances q'dí qa:-t̯'á:-d=a,
Frances good biting-sense-IPFV.SG=IMM
'It is tasting good to Frances.'

The pitch and rhythm of the second were an echo of the first (see Figure 7).

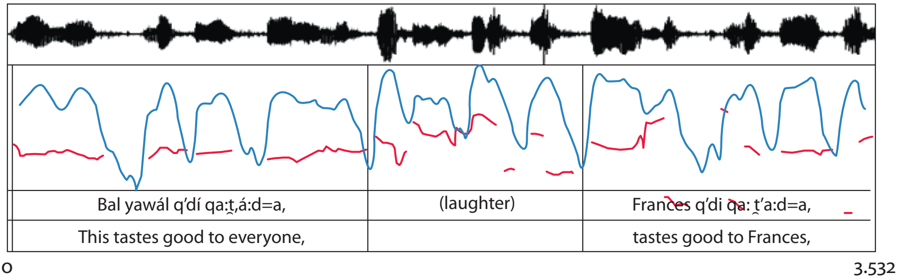


Figure 7. Statement + elaboration

The prosody of both of these constructions conveys meaning in ways the syntax does not.

4. Subjects, objects, and topicalization

Basic constituent order in Central Pomo is predicate-final. Not surprisingly, sentences consisting of a full lexical subject, a full lexical object, and a predicate with substantial content are not common. Speakers tend to introduce one new participant at a time, then not mention it again overtly if it is a continuing topic. SV and OV order can be seen in two clauses from (1), repeated here as (14).

(14) S V
Kíiki *bá’du-w,*
older.brother marry-PFV
‘my older brother married,’
O V
Maggie *ši* *bá’du-č’* *’e.*
Maggie name one-marry-SML.PFV COP
‘married a woman by the name of Maggie.’

Each of the two simple clauses shows a small pitch reset, then a relatively steady declination in pitch (see Figure 8).

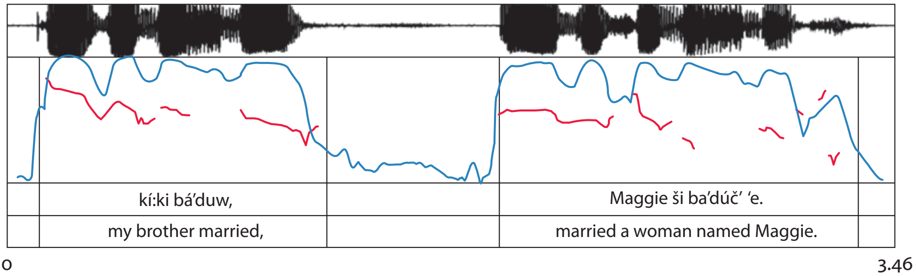


Figure 8. SV, OV, steady declination

The apparent peak on the second clause is the fricative *š*.

There is another construction which might appear to be similar in terms of grammar, with arguments preceding the verb, as in (15).

(15) S O V
Eileen *t̪’a:* *t̪ʰédu:* *hínt̪il* *čanó-:n* *t̪ʰí-n.*
Eileen guess much Indian talk-IPFV.SG be.not-IPFV.SG
‘I guess Eileen doesn’t talk Indian much.’

Here the prosody is quite different, however. The initial nominal *Eileen* ended with a partial fall then was followed by a significant pause, 1.0337 seconds, before a full pitch reset (see Figure 9). This is a different construction, a topicalization or topic shift construction, whereby speakers signal a shift to a different, usually accessible topic.

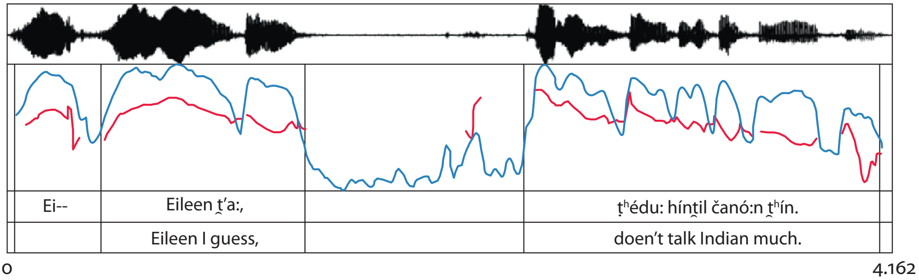


Figure 9. Topicalization

The new topic is not usually brand new. It may be accessible from previous mention in the discourse, or association with a referent that is given or semi-active in the minds of listeners. The construction has a distinctive prosodic pattern. The topicalized element precedes the nuclear clause, then is followed by a pause and a pitch reset on the nuclear clause. The initial element may be followed by one or more enclitics as here, or not.

The example in (15) above is from a conversation about a trip to the Coast to talk with other Central Pomo speakers. The distribution over Intonation Units is shown in (16), with a free translation of the immediately preceding discussion for context.

(16) They really speak a different language on the Coast;
the words are different.
It sounds different.
Not like ours.
But I understand what they say, what they want to say.
Our conversation over there must have been pretty bad.
*Ei– **Eileen** t̯'aa,*
'**Eileen** I guess,'
t̯ʰédu: *hínt̯il* *čanó-:n* *t̯ʰí-n.*
much Indian talk-IPFV.SG be.not-IPFV.SG
'doesn't talk Indian much.'

The new topic, Eileen, was accessible from earlier mention of her and the fact that everyone knew she was one of the Coast speakers. The same construction was seen earlier in (9): 'Water you know, they keep in big jugs like this'. That sentence also showed the characteristic prosodic profile of a topic shift. The topicalized element 'water' was followed by a pause of 0.8266 seconds and full pitch reset on the nuclear clause. The idea of water was not brand new at that point: It was semi-active in speakers' minds from previous discussion about finding something to drink. Even in topicalization constructions prosody can be a matter of degree, reflecting subtle distinctions. In line 27 of (17), the song was topicalized.

(17) 24 From then on they say,
25 she began a song.
26 Sang a song.
27 *Mu:l kʰé 'el 'ma kí:ki*
that song the FACT older.brother
čanó-hduwa:dan 'e.
sing.SG-FREQ.IPFV.SG COP
'My brother would sing that song too.'

Here, however, there was only a minor pitch reset on the nuclear clause, and no pause (see Figure 10). (The nuclear clause began with a voiceless stop: *kí:ki*.) The topic of the immediately preceding sentences was the woman, but the song had been mentioned in both, so it was highly accessible.

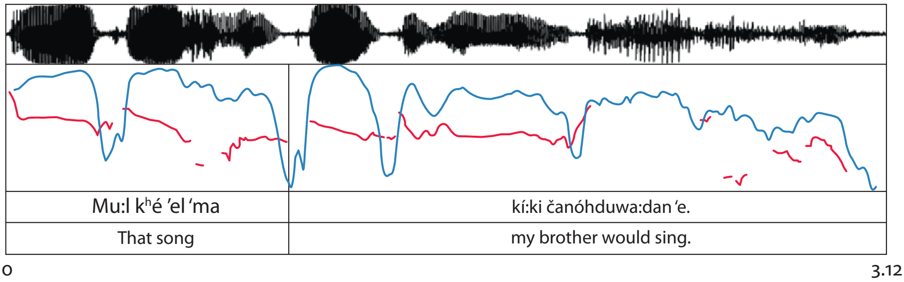


Figure 10. Topicalization

5. Clause linking

Central Pomo has a highly grammaticalized system of enclitics and suffixes that mark relations among clauses. Loosely linked clauses, viewed as separate but related events or states, are marked by one of a set of DIFFERENT enclitics on one of the clauses. Tightly linked clauses, viewed as components of a single event or state, are marked by a series of SAME suffixes on the verb of one of the clauses. These typically share one or more participants, time, and location. The markers additionally distinguish Realis from Irrealis situations, and the Realis markers further distinguish Simultaneous from Sequential events, as shown in (18).

(18) Central Pomo Clause Linkers

		SAME	DIFFERENT
REALIS			
	SIMULTANEOUS ‘while, when, whenever’	*-in*	*=da*
	SEQUENTIAL ‘and then, when’	*-ba*	*=li*
	IRREALIS ‘and, when, if’	*-hi*	*=hla*

An example of their use is in (19). The first clause ends with the DIFFERENT SIMULTANEOUS enclitic =*da*, and the second with the SAME SIMULTANEOUS suffix -*n*.

(19) *Sí:n ʼin ʼe mu:l kʼú:-baya qʼdí*
how it.is COP that child-man good
*híč-a:q-a:-w=**da**,*
say.DISTR-PASS.PFV=**DIFF.SIM**
‘How is it that they say that boy is good **and** (DIFFERENT)
*qʰá-čahá škʼé qʼóhda:du-**n** me:n má:*
water-strong only drink.HAB-**SAME.SIM** such things
ba:séṭʼay yhé-:n?
bad.DISTR do.IPFV.SG
he’s just drinking **and** (SAME) doing bad things?’

Grammar and prosody often work in concert in such constructions. Clauses expressing what are packaged grammatically as DIFFERENT but related events are usually separated prosodically. Those packaged as components of the SAME event are usually more closely integrated prosodically. The first clause in (19) ‘How is it that they say that boy is good’, marked with the DIFFERENT event enclitic =*da*, ended with a final terminal fall in pitch and was separated from the following clause by a pause of 0.4389 seconds. The next clause began with a partial pitch reset. That clause ‘he’s just drinking’, marked with the SAME event verb suffix -*n*, was not separated prosodically at all from the clause after it, ‘and doing bad things’: It was part of the same Intonation Unit (see Figure 11).

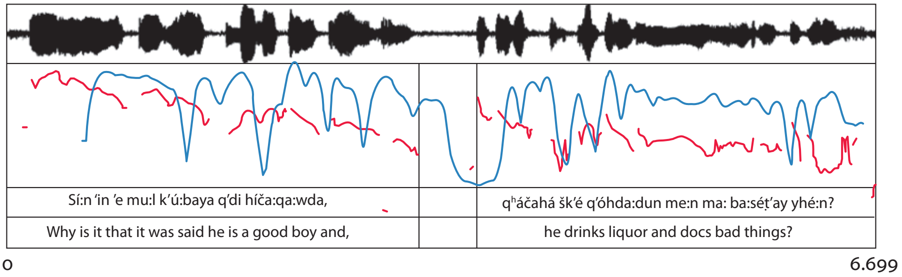


Figure 11. DIFFERENT versus SAME events

The pattern is pervasive. A sentence with the IRREALIS DIFFERENT marker =*hla* is in (20).

(20) *Čá:ʔ=ya ba: qa- ba:ʔá qó=be=**hla**,*
person=TOP someone food to=carry=**IRR.DIFF**
'If a person brings you food,'
mu:l nísmač' hní:-ka-m.
that turn.away avoid-CAUS=IMPER
'don't turn away from it.'

There was a terminal fall at the end of the first clause, a pause of 0.4557 seconds, then a pitch reset at the beginning of the second clause (see Figure 12).

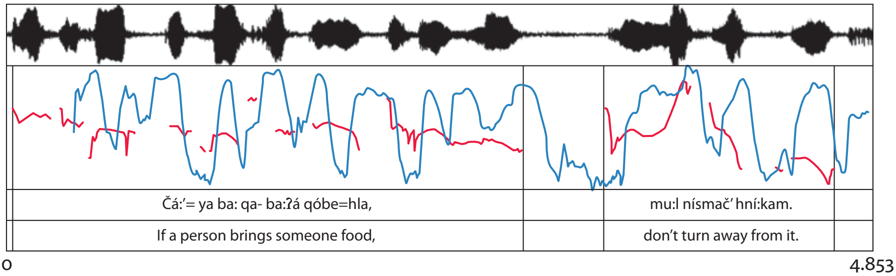


Figure 12. IRREALIS DIFFERENT events: =*hla*

A similar relation between prosody and grammar can be seen in (21), where the clauses were linked by the SAME SIMULTANEOUS suffix -*in*.

(21) *Dú:-ṯay yačól qa-ná:n-muč'-**in***
other-DISTR 3PL.OBL biting-outdo-REFL-**SAME.SIM**
'He was trying to outdo the others **while**'
qa-wá-an.
biting-go-IPFV.SG
'eating.'

There was no pause between the two clauses, and no pitch reset on the second clause (see Figure 13). They constituted a single Intonation Unit and prosodic sentence. (The apparent pitch spike is from the affricate *č'*. The brief silence on the spectrogram is the closure for the uvular stop *q*.)

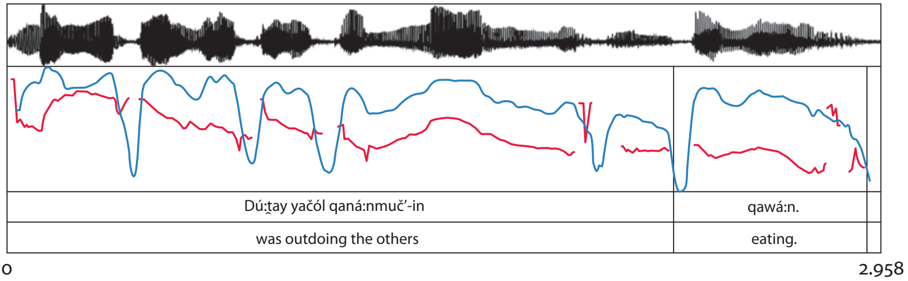


Figure 13. REALIS SAME SIMULTANEOUS event: *-in*

A similar relationship can be seen in Example (22). This sentence consists of two clauses linked with the REALIS SAME SEQUENTIAL suffix *ba* ‘and then’. There was no terminal fall in pitch after the first clause, no pause, and no pitch reset on the second clause (see Figure 14).

(22) *Mé:n ts’íba ’dóma mu:l mú:t̯u da:čé-**ba***
so then HRS that 3SG.PAT grab-**SAME.SEQ**
‘So then she grabbed hold of her **and then**’
mú:tu ya:wál yhé-:n.
3SG.PAT everything do-IPFV
‘did everything to her.’

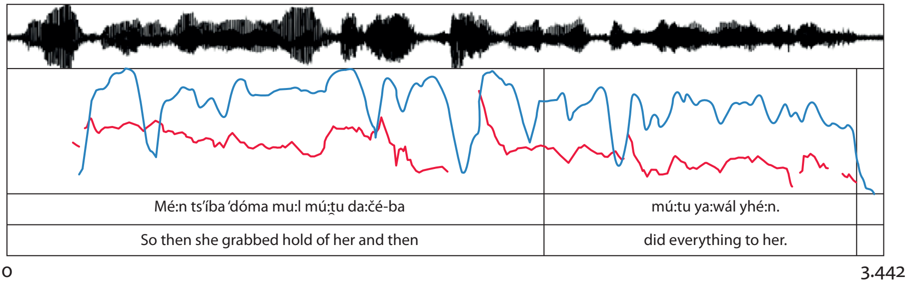


Figure 14. REALIS SAME SEQUENTIAL *-ba*

But the prosody does not always match the grammar. The sentence in (23) consists of two clauses linked by the DIFFERENT SIMULTANEOUS enclitic =*da*, but the whole was pronounced as a single Intonation Unit. There was no terminal fall in pitch after the first clause, no pause, and no pitch reset at the beginning of the second clause (see Figure 15).

(23) *Šíyal čí-w=**da** yá ’el ’úda:w*
evening become-PFV=**DIFF.SIM** wind the very
yá-č’i-dan.
blow-INCH-IPFV.SG
‘Towards evening it gets pretty windy.’

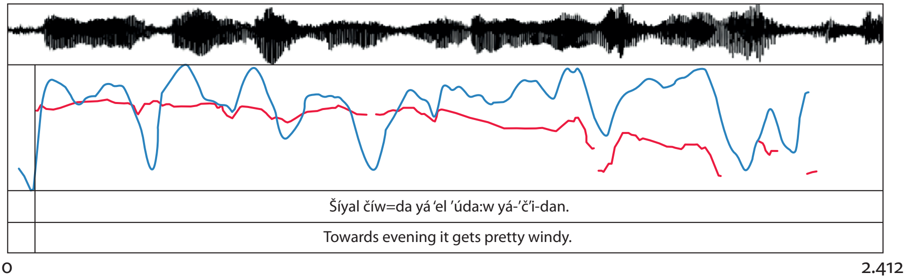


Figure 15. REALIS DIFFERENT SIMULTANEOUS events: =*da*

The sentence is literally ‘When it becomes evening, the wind starts to blow a lot’, but the speaker later translated it as ‘Towards evening it gets pretty windy.’ The first clause supplied a temporal setting rather than significant news on its own. As some friends were discussing the health of an acquaintance, one commented that the only thing that might help him would be surgery. She then relayed what the man’s wife had told him about that possibility, shown here in (24). Her warning consisted of two clauses linked by the IRREALIS DIFFERENT enclitic =*hla*.

(24) *Smá miṭí:-č-ka-ya=***hla**
sleep lie-INCH-CAUS-PASS=**IRR.DIFF**
‘If they put you to sleep’
mt̯o q’ó’t̯i madúma-č’=kʰe tʰí-n.
2SG.PAT at.all awake-INCH.PFV=FUT not-IPFV.SG
‘you’re not going to wake up at all.’

Though the clauses were linked with the DIFFERENT Event enclitic, the full sentence was pronounced as a single Intonation Unit. It showed a coherent overall declination in pitch, no pause between the clauses, and no pitch reset on the second clause (see Figure 16).

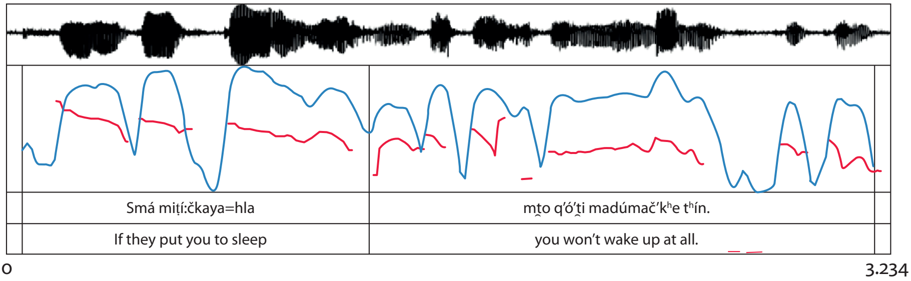


Figure 16. IRREALIS DIFFERENT events =*hla*

The prosody reflects the status of the information in the discourse. The initial clause conveyed information accessible from the discussion of surgery. In animated conversation, relations between grammar and prosody can be complex. A group was discussing the route we had taken to get to their house. Many of the points discussed so far can be seen in the excerpt in (25).

(25) EO *Béda=ht̯ow ʼe ʼma,*
here=from COP FACT

qó:=ča-:ka-w=kʰe.
hither=run.SG-CAUS=IRR
'They drove in from this side.'

FJ *Yeah.*

Kʰčé ʼmí: ʼe ya,
bridge there COP 1PL.AGT

čá-m-ma-w.
run.SG-MULT.AGT-across-PFV
'We drove over the bridge.'

Béda=ht̯ow.
here=from
Here.'

EO *ʼé:.*
'Yes.'

WL *Kʰčé ʼmí:?*
bridge there
'The bridge there?'

FJ *ʼé:,*
'Yes,'

béda=ht̯ow,
here=from

šó:=ht̯ow,
east=from

*hlá-:n-**ba** mída.*
run.PL-IPFV-**SAME.SEQ** there
'and then we came from the east.'

EO Yeah.

WL Uhuh.

The first sentence, 'They drove in from this side', was presented in two Intonation Units, separated by a pause of 0.3001 seconds. The direction 'from this side', new information, was presented in a separate Intonation Unit from the verb 'they drove here' (see Figure 17).

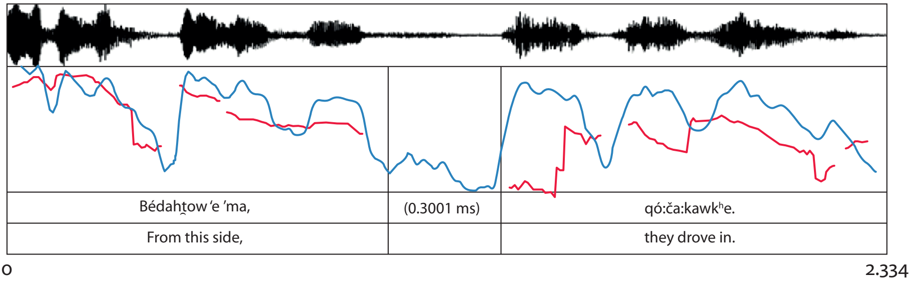


Figure 17. Two Intonation Units

In the next comment the bridge, a new idea, was also presented in its own Intonation Unit. It showed a declination in pitch, then was followed by a brief pause and a pitch reset at the beginning of the next Intonation Unit (see Figure 18).

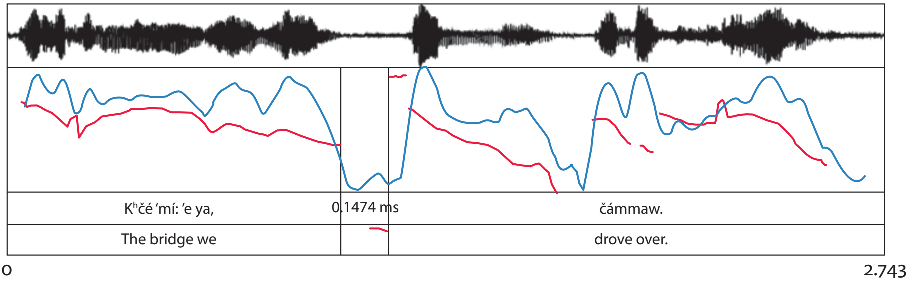


Figure 18. Two Intonation Units

Finally the clause ‘And then we came here from the east’ constituted an independent prosodic sentence, though it was dependent syntactically on ‘we drove over the bridge’, marked by the realis same sequential suffix *-ba* ‘and then’ (see Figure 19).

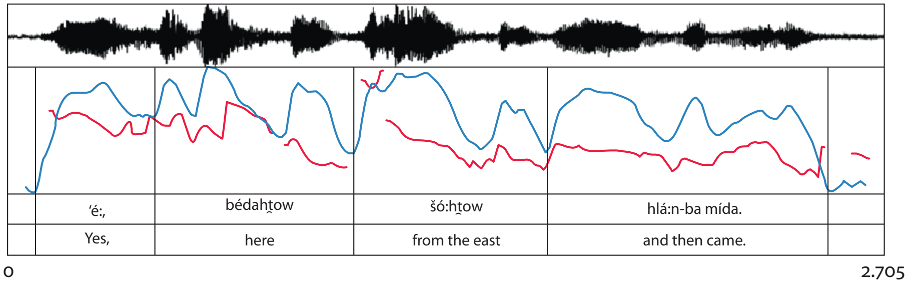


Figure 19. Dependent syntactic sentence, independent prosodic sentence

6. Conclusion

Prosodic and syntactic structures often work in concert, as seen here in examples from Central Pomo. Prosodic sentences often correspond to syntactic sentences. Such parallel patterning is common in both simple sentences and complex complement constructions. Intonation Units often consist of a syntactic constituent. But prosodic structure is not a direct reflection of syntactic structure. The two often correspond because each reflects a certain organization of ideas, but they differ in their fundamental nature. Prosody involves continua and can reflect certain subtle differences in cognitive state, discourse context, and interactive goals. Grammar (morphology and syntax) is more conventionalized and categorical.

As seen in the Central Pomo examples, each Intonation Unit introduces no more than one significant new piece of information: a participant, a time, a place, an elaboration, a whole event, etc. They may correspond to syntactic constituents, but at varying levels depending on their information status at that point in the discourse. Because prosody is not categorical, it can convey degrees of relationships among sentences, clauses, or words. Longer pauses between prosodic sentences generally signal more substantial discourse breaks. In clause-combining constructions, there is often more prosodic separation between clauses linked by enclitics marking DIFFERENT events than those linked by verbal suffixes marking components of the SAME event. The match is not perfect, however. In some cases, clauses linked by DIFFERENT enclitics are integrated prosodically because of the status of the information they display. And prosody can mark varying degrees of cohesion.

Prosody can also signal constructions not expressed by sequences of words alone. The primary difference between a basic simple clause consisting of an initial lexical nominal followed by a predicate, and a topic shift construction, is prosody. The topicalized constituent ends with a fall in pitch, then is followed by a pause and a pitch reset on the nuclear clause. In Central Pomo couplet constructions, a device often used to emphasize a point, parallel prosody marks parallel content. Elaboration constructions, also characterized by parallel prosody between clauses, serve to regulate the flow of significant new information.

While morphology and syntax can specify detailed relationships among ideas, prosody can be more directly sensitive to the cognitive state of interlocutors at the moment, and to degrees of routinization of recurring sequences.

Abbreviations

AGT	Grammatical agent
CAUS	Causative
CONTR	Contrastive
COP	Copula
DIFF	Different event
DISTR	Distributive
FUT	Future
FACT	Factual evidential
HRS	Hearsay
IMM	Immediate
IMPER	Imperative
IPFV	Imperfective
IRR	Irrealis
MULT	Multiple
PASS	Passive
PAT	Grammatical patient
PFV	Perfective
PL	Plural
REFL	Reflexive
SEQ	Sequential
SIM	Simultaneous
SML	Semelfactive

References

Chafe, W. (1979). The flow of thought and the flow of language. In T. Givón (Ed.), *Discourse and syntax* (pp. 159–181). New York, NY: Academic Press.

Chafe, W. (1984). *Cognitive constraints on information flow. Berkeley cognitive science Report No. 26*. Berkeley, CA: Institute of Cognitive Studies, University of California.

Chafe, W. (1987). Cognitive constraints on information flow. In R. Tomlin (Ed.), *Coherence and grounding in discourse* (pp. 21–51). Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.11.03cha

Chafe, W. (1988). Linking intonation units in spoken English. In J. Haiman & S. A. Thompson (Eds.), *Clause combining in grammar and discourse* (pp. 1–27). Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.18.03cha

Chafe, W. (1992). Information flow. In W. Bright (Ed.), *Oxford international encyclopedia of linguistics* (Vol. 2, pp. 2215–2218). New York, NY: Oxford University Press.

Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago, IL: The University of Chicago Press.

CHAPTER 4

Syntactic and prosodic segmentation in spoken French

Jeanne-Marie Debaisieux and Philippe Martin
Université Paris 3 Sorbonne Nouvelle LaTTiCe UMR 8094 /
Université Paris 7 Denis Diderot UFRL, LLF UMR 7110

This chapter presents first of all the analytical framework adopted for the syntactical study of spoken French productions. In line with the work of the pronominal approach, the framework postulates that three components are involved in the constitution of utterances. Two syntactical components, micro- and macrosyntax, and a prosodic component interact independently in the constitution of units. The second part of the chapter presents the application of this framework to the analysis of a conversation excerpt and an excerpt from a monologue.

Keywords: macrosyntax, microsyntax, prosodic component, pronominal approach, spoken French

1. Introduction

The concept of the sentence as the main syntactic unit cannot satisfy those who work on spontaneous spoken corpora. As pointed out by Mithun (2008),

> [i]f our syntactic analyses are based uniquely on single sentences constructed or elicited in isolation, we may miss some of the subtleties of the syntactic structures we are trying to understand, even in languages with literary traditions. (p. 72)

On theoretical grounds, Halliday (1989) underlined the need to define new units in order to overcome the shortcomings of a sentence-based approach:

> The clause complex will be the only grammatical unit which we shall recognize above the clause. Hence there will be no need to bring in the term 'sentence' as a distinct grammatical category. We can use it simply to refer to the orthographic unit that is contained between two stops. (p. 193)

https://doi.org/10.1075/scl.94.04deb
Additional files available at http://doi.org/10.1075/scl.94.04deb/audio

In pioneering work on spoken French, Blanche-Benveniste and Jeanjean (1987) noticed that "in speech, it is impossible to segment something corresponding to the notion of sentence in writing" (p. 89).[1]

Example (1a), from an interview with a writer on a French cultural channel, will illustrate this point. The speaker is talking about the problems of integration in a radio interview.

(1) a. *Moi je suis relativement optimiste parce que je pense que actuellement nous avons une génération qui est la pire celle des beurs mais ils vont avoir des enfants à leur tour et la troisième génération ils seront Français*
'I am relatively optimistic because I think we have now a generation that is the worst, the children of North African immigrants but they will have children of their own and the third generation they will be French'[2]
[source: radio broadcast]

If one wishes to respect the purely syntactic division into standard French sentences, one must first isolate the following segment which presents a complex sentence structure which includes a main verbal construction and a subordinate construction introduced by *parce que* 'because'. But the result appears semantically absurd in (1b):

(1) b. ?? *Moi je suis relativement optimiste parce que je pense que actuellement nous avons une génération qui est la pire celle des beurs*
??'I am relatively optimistic because I think we have now a generation that is the worst the one of the children of North African immigrants'

To maintain the semantic coherence of the remarks made by the speaker, it is necessary to include both the *but*-clause and the *and*-clause under the scope of the conjunction *parce que* 'because' since if we remove these two clauses, the remaining part cannot be properly interpreted given that one cannot be optimistic about a future involving the worst generation of young people.

This contradiction cannot be explained by the speaker's poor knowledge of the language or by the communicative situation since the speaker is a reputed writer in a cultural broadcast. To understand (1) we must give up the idea of segmenting the utterance into sentences. We must consider that this utterance comprises the entire sequence in which the speaker develops a complex argumentation to justify his optimism. Thus, it is necessary to include under the scope of the conjunction *parce*

[1] "Une des notions qui saute c'est celle de phrase ; impossible de découper dans le parlé quelque chose qui corresponde à la notion de phrase pour l'écrit" (Blanche-Benveniste & Jeanjean, 1987, p. 89).

[2] France Culture Radio, 1988: <https://www.franceculture.fr/>

que a whole paragraph composed of three syntactically independent clauses even if this goes against the normative rules of how sentences are constructed in French.

In fact, in spontaneous data, syntactic units are not defined *a priori,* they have to be established from the data by explicit descriptive steps. The first step is to segment the raw text into units. The second step, depending on the syntactic purpose, is to determine the internal composition of the units, and define their possible combinations, that is, what kind of relation we can observe between units. For these latter steps, we propose to distinguish two kinds of syntactic relations: micro- and the macrosyntactic dependency.

The first section of this paper is theoretically oriented, and sets out our linguistic framework: the *Approche Pronominale* (Blanche-Benveniste et al., 1990, Blanche-Benveniste & Martin, 2010; Debaisieux, 2013; Deulofeu, 2003). This analytic framework combines two syntactic subcomponents, microsyntax and macrosyntax, and a prosodic component. These three components are independent and interact in different ways in the construction of utterances.

The second section of this chapter is descriptive: We will analyze from this approach the various units unearthed in a conversation extract and then in a monologue extract from two corpora of transcribed spontaneous spoken French: the TCOF Corpus <www.cnrtl.fr/corpus/tcof/>[3] collected in the French city of Nancy, and the French part of C-ORAL-ROM <http://www.lllf.uam.es/ING/Coralrom.html>.[4] The conversation involves three friends who are discussing their holiday plans and who tell a few anecdotes. They tell how one of their friends mixed very quickly with strangers during a party. The monologue is a story of cruising and scuba diving. We wanted to work on these two extracts because we do not want to restrain the analysis to conversations. We will see, moreover, that there are, of course, similar units in both texts. The examples are referenced according to their

[3] The Treatment of Oral Corpus in French (TCOF) Corpus was born from the desire to preserve oral corpora collected in the 80s-90s for personal research purposes. It portrays recordings of adult-child interactions (children up to 7 years old) and of interactions between adults. The recordings are of various durations: from 5 to 45 minutes or more. The corpus was the first one in France to have text-sound alignment, additionally it makes available the transcriptions and sound files.

[4] C-ORAL-ROM is a multilingual corpus of spoken romance languages: French, Italian, Portuguese and Spanish. The project was funded by the EU within the V Framework Programme (IST-2000-26228) and the consortium comprises nine partners coordinated by the University of Florence. The most significant feature of C-ORAL-ROM is the spontaneity of texts: they were recorded in real context and without a script. Each subcorpus is made up of 300.000 words, with the same textual distribution to guarantee comparability and representativity. The resource is presented in different formats: an orthographic transcription, an XML tagged version and the text-sound alignment.

origin: TCOF for the Nancy corpus, FFAMCV11 for the conversation, FFAMNN01 for the monologue in C-ORAL-ROM. The purpose of this article is to show how, by using these three components and their complementarity, we can give a natural account of the syntax of spoken language productions.

2. Linguistic framework

2.1 From text units to discourse units

Text units correspond to what is generally meant by "utterance", that is, segments of discourse that are syntactically independent and prosodically and semantically autonomous. The notion of utterance is also sometimes used to designate a segment made up of non-clausal material because it signals that a non-clausal syntactic frame (e.g., the adverb *forward*) "unexpectedly" plays the role that is usually assigned to a main clause, because it carries illocutionary force. This is the case for the adverb *forward* in an exclamation such as *forward*!, in which instead of it being a mere constituent integrated in a clause as in *I am moving forward* (Culicover & Jackendoff, 2005, p. 236), it actually is an illocutionary unit. But if the text unit is clause-like, linguists no longer feel that it is useful to distinguish between a clause and an utterance. Yet the difference can still be related to a clear formal property: A clause is a syntactic frame integrated into a construction without an independent prosodic contour, whereas an utterance is a syntactic frame endowed with an autonomous prosodic contour. For descriptive frameworks, it is not at all surprising that text units or utterances may appear to be built upon clausal as well as non-clausal frames. Such evidence leads descriptive linguists to replace the definition of the sentence as the maximal syntactic unit by an alternative one: (Almost) any word or phrase or combination of phrases can form a syntactic frame for building a text unit if endowed with illocutionary force by the appropriate prosodic contour. A main clause is just one type of text unit among possible ones, the one that is based on a construction headed by a finite verb. However, one more step should be taken to capture the syntactic structure of spontaneous discourse: We must go beyond text units and move on to discourse units.

Discourse is usually reduced to a concatenation of utterances, which boils down to saying that blocks of discourse, in other words utterances, are necessarily formed by means of syntactic frames, clausal or non-clausal. But the actual units that speakers use to convey messages to their addressees go far beyond these forms. Addressees also accept messages without syntactic frames. Some of them are phonetic segments, such as interjections or onomatopoeias. Admittedly, they have phonetic substance, but they are not integrated into the grammatical system of the language. Other messages have no phonetic content and consist of what we may call communicative behaviors: facial expressions, gestures. But discourse units are also made with pieces of meaning derived by inference from what has been said by the speaker. This is the case in Example (2) from a telephone conversation:

(2) Odile (L2) has taken L1's kids for a walk in the Pépinière (a park).[5] They arranged to meet later in the evening. But Odile calls to pick up the kids sooner.
L2 *Allô c'est Odile*
'Hello it is Odile'
L1 *Oui*
'yes'
L2 *Parce qu'il fait froid à la Pépinière*
'Because it is cold in the Pépinière'
L1 *Il faut venir vous chercher tout de suite ?*
'Shall I come and fetch you right away?'
L2 *Oui*
'Yes' [source: TCOF]

Speaker L2 begins her second turn with a subordinate clause. The only candidate for a main clause is the *c'est Odile* of the first turn. Semantically, it would be absurd to compositionally combine the two propositional contents *c'est Odile parce qu'il fait froid à la Pépinière* 'it is Odile because it is cold in the Pépinière'.

Moreover, the semantic anchor of the subordinate clause is neither the propositional content nor the speech act force of any utterance in the context. We have to consider that the anchor is the general scenario of making a phone call, suggested both by the practical action of calling and the accompanying words.

Berrendonner (2003), Blanche-Benveniste et al. (1990), Deulofeu (2008), and Debaisieux (2013, 2016) claim that it is possible and necessary to capture the combinatorial regularities of discourse units in a separate component of the linguistic description, *macrosyntax*. These units have some specific features that cannot be strictly described by grammatical relations. As pointed out by Blanche-Benveniste (2010), "Utterances produced by speakers include composite materials of syntax, prosody, semantics, and pragmatics, as well as a range of speech routines" (p. 159).[6] However, they exhibit some regularities, for example in their distributional constraints and degrees of enunciative and illocutionary autonomy. It is therefore necessary to articulate these regularities with the rules of syntax in the narrow sense (micro-syntax) in order to describe properly the way in which messages are processed.

[5] The two speakers are represented by numbers: L1 and L2.

[6] "Les énoncés produits par les locuteurs comportent des matériaux composites de syntaxe, de prosodie, de sémantique, de pragmatique, ainsi que tout un ensemble de routines de discours". (Blanche-Benveniste, 2010, p. 159)

These two subcomponents, micro- and macrosyntax originate from the need to describe real structures, in particular in spontaneous spoken corpora. Let us examine Example (3) and Example (4), adapted from a well-known example given by Culioli (1983):

(3) *Le guidon du vélo de mon frère est cassé*
'The handlebar of my brother's bike is broken'

(4) *Mon frère son vélo le guidon il est cassé*
'My brother his bike the handlebar it is broken'

Example (3) is a canonical French sentence that we could analyze in traditional grammatical terms: We have a verbal construction based on the grammatical relation between a governor, the verbal construction *être cassé* 'to be broken', and its dependent: here a complex nominal phrase in which the dependency is grammatically marked. To interpret this sentence, we can just sum up the different meanings of the parts, whether lexical or grammatical. This sentence illustrates Frege's Principle of compositionality: "The meaning of a whole is a function of the meanings of the parts and the way they are syntactically combined" (Partee, 1995, p. 313).

Example (4) is very different. It is impossible to describe this unit using only the traditional syntactic component: Some parts are not grammatically linked. The three nominal phrases *mon frère* 'my brother', *son velo* 'his bike', *le guidon* 'the handlebar', must be seen as autonomous. They are not in a dependency relation with the verb and there are many factors – lexicon, grammar but also prosody and inferences – which contribute to the interpretation of the whole as a unit. This example should be analyzed by extending the syntactic component in order to account for the way in which grammatical constructions are used to build utterances and discourse. As Blanche-Benveniste (2010) remarked:

> Since these organizations cannot be organized solely by the syntax of grammatical categories, several recent studies concur in situating them at a more encompassing level of macrosyntax. (p. 159)[7]

We therefore base the architecture of the formal component on two syntactic subcomponents (micro- and macrosyntax) and a prosodic component. The macrosyntactic subcomponent is responsible for combinations of syntactic units at the discourse level. All the components are interrelated through distinct sets of interface rules.

[7] "Comme ces organisations ne peuvent pas être organisées uniquement par la syntaxe des catégories grammaticales, plusieurs études récentes se sont accordées pour les situer à un niveau plus englobant de macrosyntaxe". (Blanche-Benveniste, 2010, p. 159)

2.2 The syntactic subcomponent

The basic assumption is that syntactic units combine in two different ways. The first one is classically achieved by a recursive use of dependency relationships between a syntactic head and its dependents. This amounts to building larger constructions from smaller ones. The second one uses paratactic links between utterances built upon constructions. Utterances are the building blocks of discourse and they combine into discourse patterns.

In our specific framework, *Approche Pronominale*, the two syntactic subcomponents are organized as follows: In the microsyntactic subcomponent, constituents are linked by government relationships, that is, dependency on a governor category (N, V, P…), and constructions are based on projections of categories. The grammatical dependency relation can be demonstrated by means of a set of syntactic tests among which suppression, pronominalization, clefting, and embedding. Different frameworks rely more on some of these tests. We consider that suppression and embedding are not reliable tests because they can diagnose grammatical as well as discourse dependency. For example, compare the following two statements ((5a), (6a)) in which *puisque* 'since' and *parce que* 'because', which are generally considered as introducing a subordinate construction, are used. The two constructions react differently to tests, as can be appreciated in the following discussion:

(5) a. *Paul est arrivé de bonne heure parce qu'il devait parler le premier*
'Paul arrived early because he had to speak first'

(6) a. *Paul est arrivé de bonne heure puisque tu l'avais prévenu*
'Paul arrived early since you had warned him'

a. The clause with *parce que* (5b) can be clefted whereas the *puisque* clause (6b) cannot:

(5) b. ***C'est*** *parce qu'il devait parler le premier* ***que*** *Paul est arrivé de bonne heure*
'It was because he had to speak first that Paul arrived early'

(6) b. ****C'est*** *puisque tu l'avais prévenu* ***que*** *Paul est arrivé de bonne heure*
*'It's since you warned him that Paul arrived early'

b. The clause with *parce que* (5c) can be an answer to a *why* question whereas the *puisque* clause (6c) cannot:

(5) c. *Pourquoi Paul est-il arrivé de bonne heure ? Parce qu'il devait parler le premier*
'Why did Paul arrive early? Because he had to speak first'

(6) c. **Pourquoi Paul est-il arrivé de bonne heure? Puisque tu l'avais prévenu*
*'Why did Paul arrive early? Since you warned him'

c. The clause with *parce que* (5d) can be modified by a scope adverbial whereas the *puisque* clause (6d) does not accept this modification:

(5) d. *Paul est arrivé de bonne heure principalement parce qu'il devait parler le premier*
'Paul arrived early mainly because he had to speak first'

(6) d. **Paul est arrivé de bonne heure* ***principalement*** *puisque tu l'avais prévenu*
*'Paul arrived early mainly since you warned him'

Therefore *puisque tu l'avais prévenu* 'since you warned him' will be analyzed as a discourse unit linked by a discourse relationship and not linked grammatically as an adjunct subordinate clause. These examples show that a correlation cannot be postulated between syntactic and morphological levels.

At the microsyntactic level syntactic frames (phrases and clauses) combine into larger frames in accordance with grammatical dependency rules. At the macrosyntactic level, discourse units, that is utterances (frames + prosodic contours) and communicative behaviors combine according to the regularities of what Mithun (2005) calls pragmatic dependency. Therefore, it is important to emphasize the difference between grammatical and discursive dependency. By splitting syntax into micro and macro subcomponents, we systematize the distinction between syntactic and pragmatic dependency put forward in Mithun (2005):

> […] markers [of syntactic dependency] are being used to signal pragmatic dependency among larger elements in discourse. The markers of dependency serve several recurring functions in discourse. The Yup'ik Participial and Barbareno nominalized sentences contribute background, descriptive, subsidiary, explanatory, or evaluative information, information that does not move narrative forward. The Yup'ik Subordinative and the Hualapai switch-reference markers signal textual cohesion, marking statements that together compose a larger discourse unit. (p. 89)

The macrosyntactic subcomponent accounts for how constructions can form discourse units (utterances) and how discourse units combine to build discourse patterns. In this subcomponent, the constituents are not directly linked by grammatical dependency relations. The units are defined by means of illocutionary or communicative properties and are characterized by autonomous prosodic contours related to modal and illocutionary features.

Within the macrosyntactic subcomponent, we distinguish two types of units: free units or the "Nucleus", and dependent units or the "Satellite". The Nucleus can stand by itself as an autonomous free-standing utterance. As for its internal composition, it is by default a construction endowed with an illocutionary force. One way of testing whether a constituent is endowed with illocutionary force is to check that it can carry "utterance modalities", which is evidenced for a Nucleus built on a clause by the fact that the full range of sentence types is possible in its position, as in (7):

(7) *Il est arrivé / Est-il arrivé ? / Arrive !*
'He has arrived' / 'Has he arrived?' / 'Come here!'

This property may be interpreted in terms of contrasting illocutionary forces (assertion, question, order) and coded by the feature [+illoc] (Verstraete, 2007). As Deulofeu (2013) points out, the [+illoc] feature is not attributed on the basis of a mere semantic intuition but as the result of the observation that one can build significant oppositions of meaning between substitutable forms (Deulofeu, 2013, p. 486).[8] All syntactic types of Nuclei are further characterized by the fact that they can bear a range of terminal prosodic contours contributing the illocutionary force (coded here by punctuation marks) as in (8) to (11):

(8) *Il est arrivé. / Il est arrivé ? / Il est arrivé !*
'He has arrived' / 'Has he arrived?' / 'He has arrived!'

(9) *Ce que c'est beau !*
'How beautiful it is!'

(10) *Quand je pense qu'il devait venir !*
'To think he was coming!'

(11) *Est-ce qu'il est arrivé ?*
'Has he arrived?'

To sum up, the Nucleus is essential for the processing of discourse: It may constitute a complete message, acknowledged as such by the addressee, and can fulfil various speech functions (Verstraete, 2007) depending on its specific terminal prosodic contour. This central unit may be accompanied by one or several Satellites, which can be considered as discursively or pragmatically dependent on the Nucleus as they cannot form a free-standing message by themselves but need to be grouped with a Nucleus to be properly interpreted. As for internal composition, the Satellite bears the feature [–illoc] irrespective of whether it is realized before or after the Nucleus. The [–illoc] feature codes the fact that the construction displays a non-terminal prosodic contour and that no sentence type variation is possible, as in (12a) and (12b):

(12) a. *Comme il était là je ne suis pas venu*
'As he was there I didn't come'

(12) b. **Comme est-ce qu'il était là, je suis pas venu*
*'As was he there I didn't come'

[8] This ensures that our macrosyntactic units have the full status of "signs".

On prosodic grounds, the Satellite bears a non-terminal contour, which is not related to the illocution domain, but displays a merely cohesive function of grouping the Satellite and the Nucleus. Beside their differences in prosodic and speech act status, Nuclei and Satellites differ in terms of other formal features. The Nucleus does not have a fixed microsyntactic composition: It is compatible with main clause phenomena. Furthermore, it can, for instance, take the form of a concatenation of clauses forming a long stretch of discourse as in (2) in which the conjunction has scope over what could be considered a paragraph in spoken language.[9] By contrast, main clause phenomena are excluded in Satellites and the scope of the conjunction is only local. Satellites can be placed either before or after the Nucleus, but some Satellites have topological constraints: For example, the *comme* clause in (12) can never be placed after the Nucleus. Furthermore, some particular constructions are "specialized" for a specific macrosyntactic function: *Il a beau* + Vinf 'although', can be only a Satellite, *tant pis* 'too bad', only a Nucleus.

2.2.1 *Combination of microsyntactic units*

Constructions are based on projections of categories and a combination of units forms larger constructions by the recursive use of dependency relations based on a category as governor. Tests show that the construction falls within a restricted paradigm of syntactic forms triggered by the subcategorization frame of the governor.

2.2.2 *Combination of macrosyntactic units*

The macro-units usually combine by concatenation into different discourse patterns. The first pattern can be called "extended utterance". It involves only one Nucleus and a variable number of Satellites. It can be linearized into several topological configurations in which three subtypes of Satellite can be distinguished, according to their position before, after or inside the Nucleus: Pre-Nucleus, Intra-Nucleus, Post-Nucleus; as in (13).

(13) [*Comme il faisait nuit*] [*il est arrivé tard*] [*si tu veux*] [*à cette réunion*]
Satellite (Pre) ***Nucleus part 1*** ***Satellite (Intra)*** ***Nucleus part 2***
[*le type*]
Satellite (Post)
'[As it was dark] [he came late] [you see] [to this meeting] [the guy]'

As an example of a second type of discourse pattern we can mention the concatenation of "extended utterances" involving a conjunction as discourse connective, as in (14).

[9] This amounts to saying that a Nucleus must contain at least one subpart bearing a [+illoc] feature.

(14) *Paul est à la fac* *parce que* *n'oublie pas que c'est son jour de cours !*
Nucleus ***Discourse connective*** ***Nucleus***
'Paul is at the college because don't forget that this is his seminar day!'

The configuration can correspond to a textual organization in which various Nuclei are concatenated. Furthermore, macrosyntactic units, as communicative units, are by definition multi-semiological. As such they can be realized by segmental components (constructions) as well as gestural components (gestures, attitudes, and contours). They may comprise only a gestural/mimetic attitude or ostensibly imply an element of the extra-linguistic situation (Debaisieux & Deulofeu, 2001). A Nucleus can be inferred by the addressee on instruction by the speaker or directly from the context. It should be pointed out that the Nucleus is characterized by a terminal prosodic contour. Each contour is associated with a default illocutionary interpretation: assertion, question, and their variants, evidence, doubt, command, surprise (Martin, 2015). The Satellite has no terminal contour and cannot host, if clausal, a full range of sentence types: It is thus deprived of illocutionary force. To say that such and such a microsyntactic unit stands as such and such a macrosyntactic unit amounts to saying that it can be associated with such and such a prosodic contour (Cresti & Moneglia, 2010).

We will illustrate this point in the next section. Indeed, intonation plays a crucial part in distinguishing the main types of discourse units, as it seems to be the main formal counterpart of illocutionary force assignment.

2.3 Prosodic component

In order to analyze better the interactions between macrosyntax and prosody, we consider two hierarchical organizations of the sentence separately. This means that parallel to the macrosyntactic structure, we assume that another structure organizing prosodic units does exist, *a priori* completely independent from the syntactic and macrosyntactic events. This non-conventional approach provides many benefits, in particular while examining the possible correspondence between macrosyntactic and prosodic boundaries. As these boundaries do not necessarily coincide, looking for macrosyntactic boundaries marked by prosody may fail, and conversely, looking for prosodic boundaries correlated with syntax may not work either.

Considering both macrosyntactic and prosodic structures as independent, the possible configurations are: (1) Macrosyntactic boundary → no coincident prosodic boundary; (2) Prosodic boundary → no coincident macrosyntactic boundary and (3) Macrosyntactic boundary → coincident prosodic boundary. We will discuss here only cases belonging to the third configuration.

Our definition of the prosodic structure is somewhat similar to the one used in the Autosegmental-Metrical approach, with some differences due to the absence of lexical stress in French (in particular, there can be more than one content word in a single accentual phrase, depending on the speech rate; see Martin, 2018). Accentual phrases (AP) contain words of any category with one non-emphatic stressed syllable, placed on the last AP syllable. Furthermore, the tone boundaries are not described by high and low levels, but by melodic contours, with the following definitions:

a. C0: conclusive declarative contour, noted ↓;
b. C0i: the conclusive declarative contour C0 in its implicative/exclamative variant;
c. C0c: the conclusive declarative contour C0 in its imperative variant;
d. C0n: the melodic contour ending a prosodic Post-Nucleus, noted ←;
e. C1: intonation phrase (IP) boundary contour (*continuation majeure*, 'major continuation'), noted ↗;
f. C2: intermediate phrase (ip) boundary contour (*continuation mineure* 'minor continuation'), noted ↘;
g. Cn: neutralized contour, ending stress groups composing ip, noted →;
h. Ci: the terminal interrogative contour, noted ↑.

The melodic descriptions of these contours are:

a. C0: low and falling with a reduced frequency range;
b. C0i: low and falling with a hump;
c. C0c: low and falling with a large frequency range;
d. C0n: flat and long contour, below the glissando threshold;
e. C1: rising with a large frequency range, above the glissando threshold;
f. C2: falling with a large frequency range, above the glissando threshold;
g. Cn: rise or fall with a much-reduced melodic variation, below the glissando threshold.

The glissando threshold (Rossi, 1971) separates melodic contours C1↗ rising and C2↘ falling whose variation is perceived from the neutralized contour Cn→, whose melodic variation is too restrained to be perceived. The terminal declarative C0↓ and interrogative Ci↑ contours reach the lowest and highest melodic height in the sentence, whereas C0n← is a flat contour appearing only after the terminal contour C0↓. The use of the glissando threshold gives a proper account for the principle of melodic slope contrast, where the falling contour C2↘ instantiates a dependency relation towards the rising contour C1↗ which in turn indicates a dependency relation towards the falling declarative terminal contour C0↓ (Martin, 2018).

These dependency relations "to the right" applied dynamically during the course of the sentence determine incrementally the prosodic structure by merging successively accentual phrases. Accentual phrases are sequences of syllables that contain only one – non-emphatic – stress placed on the last accent phrase syllable, that is, on the final syllable of the last word of the accent phrase. Due to the lack of lexical stress in French, any category of words, whether lexical or grammatical, can belong to an accent phrase. The only constraint pertains to the time it takes to pronounce the words contained in an accent phrase, between 250 ms and some 1250 ms (Martin, 2014). A faster speech rate involves accentual phrases with a larger number of syllables and words. A slower speech rate implies fewer syllables, down to one syllable APs.

Therefore, parallel to the macrosyntactic structure, the prosodic structure is defined as a hierarchical organization of APs. The melodic contours described above indicate this prosodic structure through dependency relations. Contrary to the text macrosyntactic structure, the prosodic structure does not have Pre-Nucleus, but only a prosodic Nucleus, which can be followed by a prosodic Post-Nucleus (realized with a flat melodic contour). The Post-Nucleus corresponds to the theme in the old theme-rheme terminology.

The only obligatory alignment between macrosyntactic and prosodic structures pertains to the right boundary of text and prosodic Nucleus: Right boundaries of both text and prosodic nuclei are necessarily aligned. We have thus two distinct analyses in macrosegments for text and intonation, as shown in the Example (15). The prosodic Nucleus [C1 C0] is aligned on the two components *le métro c'est sous terre,* and the prosodic Post-Nucleus [C0n] is aligned on the second macrosegment *c'est sous terre.*

(15) *Le métro* C1↗ *c'est sous terre* C0↓ *le métro* C0n←
'The subway is underground the subway' ("Zazie dans le métro", R. Queneau)

2.4 The interface rules between the formal components

The Interface rules between the components are free and there may be congruence or no congruence between the components. We present briefly here the interface rules between the components of our framework.

a. Morphology (categories) *versus* syntax

We have seen that the morphological analysis needs to be disconnected from the syntactic one. The type of conjunction, for example, does not determine the syntactic relations that the construction has with the context. Conjunctions can link constructions or discourse patterns. Prepositions also have these two possibilities.

b. Microsyntactic units *versus* macrosyntactic units

By default, any type of macrosyntactic unit can stand as any type of microsyntactic unit. A Nucleus can be built on the microsyntactic units, both clausal and non-clausal.

c. Microsyntactic Unit *versus* prosody

We assume that another structure of prosodic units does exist, a priori completely independent from the syntactic and macrosyntactic events. For example, a microsyntactic unit can be separated into two parts by prosody. We call this phenomenon "epexegis" (see Section 3.4).

d. Macrosyntactic unit *versus* prosody

Due to prosodic autonomy, we can distinguish three cases:

1. The macrosyntactic boundary is not congruent with the prosodic boundary: A prosodic structure may group two micro-syntactically independent constructions in a single Nucleus.
2. The prosodic boundary is not congruent with the macrosyntactic boundary: Several Nuclei with non-terminal contours form a discourse pattern of the narrative type (textual organization)
3. The macrosyntactic boundary is congruent with prosodic boundary.

The third case is the main interface rule since by default, a terminal contour marks a macrosyntactic Nucleus.

e. Syntactic units *versus* meaning

Microsyntactic constructions are interpreted compositionally: A verbal syntactic frame is interpreted as a propositional content, a nominal construction as a referential entity. Macrosyntactic configurations are by default interpreted non-compositionally. Let us compare Examples (16) and (17):

(16) *On (n')a pas pu aller jusqu'à Tiran parce que c'était trop loin d'Hourghada*
'We could not go up to Tiran because it was too far from Hurghada'
[source: FFAMNN01]

Example (16) can be analyzed as a complex sentence in which the conjunction *parce que* 'because' fulfills the standard function of introducing an adjunct clause to the main verb *go*. The conjunction establishes a semantic relation (here: cause) between the propositional contents of the governing and governed clause: The fact that Tiran 'was too far from Hurghada' is indeed the cause of the fact expressed in the first construction: 'we could not go'.

Example (17) illustrates the grouping of two discourse units into a discourse pattern. Grammatically speaking, the two constructions forming two discourse units are simply concatenated:

(17) *Ils nous prennent trop pour des poires parce que les papiers tu peux pas les changer*
‘They take us too much for suckers because the papers you cannot change them’
[source: TCOF]

Besides, the semantic relationship between the clauses cannot be interpreted as a cause-effect relation. Thus, clearly, the fact that ‘you cannot change the papers’ cannot be the cause of the fact that ‘they take you for suckers’. To get a coherent interpretation we must assume that, in the configuration case, *parce que* ‘because’ establishes a paratactic link between the assertive illocutionary force of the first clause and the fact conveyed by the second clause.

3. Descriptive issues

In this section, we present a non-exhaustive typology of the types of units found in the excerpts of the conversation and monologue. We will then describe some original configurations of a monological excerpt.

3.1 Simple utterances: Nuclei

As we saw in the first part, simple utterances may be verbal or non-verbal and belong to any grammatical category. This is illustrated by the following example in which the macrosyntactic unit, the Nucleus, is made of a single microsyntactic constituent: an adverbial in (18) and built on a nominal group in (19). These examples are marked by an assertive value. On the other hand, (20) illustrates the possibility of the Nucleus having an exclamatory value. We see that any category can carry the feature [+illoc]. Any category functioning as a Nucleus can carry the feature [+illoc].

a. Non verbal Nucleus with assertive value

(18) [*Non* C0↓]^{N} [*mais non* C0↓]^{N}
‘[No] [but no]’ [source: FFAMCV11]

(19) [*Rancart* C1↗] [*pour mardi soir* Cn→ *et tout ça* Cn→ *quoi* C0↓]^{N}
‘[Date] [for Tuesday night and all that what’s the point]’
[source: FFAMCV11]

b. Non verbal Nucleus with exclamatory value

(20) [*Enfin n’im- n’importe quoi* C0↓]^{N}
‘[Finally anything goes]’ [source: FFAMCV11]

c. Verbal Nucleus

In (21) and (22), we have perfect congruence between micro- and macrosyntax: A canonical microsyntactic unit, a simple verbal phrase, constitutes a macrosyntactic Nucleus.

(21) [*Vous avez prévu* Ci↑]^{N}
'You have planned' [source: FFAMCV11]

(22) [*Je vais à la plage* Cn→ *avec mes copines* C0↓]^{N}
'I'm going to the beach with my girlfriends' [source: FFAMCV11]

(23) also presents a canonical microsyntactic constitution: a complex verbal phrase which forms a simple Nucleus. While there is a prosodic break after the first C1, it is irrelevant for the macrosyntactic structure since it does not mark a Pre-Nucleus Satellite but has simply a cohesive function: It groups the governor in a single utterance with the adjunct according to the slope contrast principle, that is, C1 contrasts with C0.

(23) [*Il l'a larguée* C1↗ *parce qu'elle avait pris* Cn→ *des kilos* Cn→ *en trop* C0↓]^{N}
'[He dropped her because she had put on too much weight]'
[source: FFAMCV11]

There are also examples that, according to standard analyses, display structures that don't formally correspond to canonical structures. In (24) the verbal construction can be considered formally subordinate, as it is introduced by a subordinating conjunction, *si* 'if'. However, it stands as an isolated discourse unit which it is impossible to link syntactically to what precedes or follows. No main clause can be found in the context that could govern the subordinate clause which behaves here as an autonomous statement and also forms an independent prosodic unit: a Nucleus.[10]

(24) *Mais si tu savais ce que moi j'étais contente*
'But if you only knew how happy I was' [source: TCOF]

3.2 Compound utterances: Nucleus + Satellite

As noted in the first part of this paper, Nuclei are often associated with one or more Satellites to form compound utterances. The most frequent cases are built on the association of a Pre-Nucleus and a Nucleus.

[10] See Debaisieux, Martin and Deulofeu (in press) for a detailed analysis.

a. Pre-Nucleus Nucleus

(25) [*La coque* C1↗]^{S} [*le sable est à quarante mètres* C0↓]^{N}
‘[The hull] [the sand is forty meters deep]’ [souce: FFAMNN01]

(26) [*Ben là depuis deux semaines* C1↗]^{S} [*je je vais je rentre le week-end* C0↓]^{N}
‘[Well after (being) there for two weeks] [I I will I go home at weekends]’ [source: FFAMCV11]

(27) [*Mais euh son mec* C1↗]^{S} [*c’est un top-model* C0↓]^{N}
‘[But uh her boyfriend] [he is a supermodel]’ [source: FFAMCV11]

The three examples present a [Pre-Nucleus/Nucleus] structure in which the Pre-Nucleus is a simple constituent: a noun phrase in (25), a prepositional phrase in (26), and a noun phrase preceded by a conjunction in (27). They show that the semantic relationship between macro units may be compositional or not. Example (25) shows a hanging topic with a non-compositional semantic link while (26) shows a temporal adjunct and (27) a classical topic. Indeed, there is no congruence between microsyntactic units, meaning and macrosyntactic units. This is illustrated in (28) and (29).

(28) [*J’ai le le copain de ma copine* C1↗] ^{S} [*il a un petit bateau* C2↘ *dans le port de Toulon* C1↗] ^{N}
‘[I have the the boyfriend of my girlfriend] [he has a small boat in the port of Toulon]’ [source: FFAMCV11]

(29) [*Mais elle les connaît pas* C1↗]^{S} [*elle part avec eux* C1↗]^{N}
‘[But she doesn’t know them] [she left with them]’ [source: FFAMCV11]

In (28) the Satellite is a bare clause with a light verb *avoir* ‘to have’ with the function of introducing the topic *le copain de ma copine* ‘the boyfriend of my girlfriend’ whereas in (29) there is a plain bare clause in the Satellite, with a subordinate interpretation: Given that it is the case that she is not acquainted with them, she goes with them. As mentioned, it is common for a Nucleus to be preceded by several Satellites. This is shown in (30), which contains two Satellites:

(30) [*Et comme les gars voulaient pas rester en mer*]^{S1} [*ils voulaient rentrer tous les soirs*]^{S2} *euh* [*je pense qu’on a bouffé pas mal de mazout pour pour faire les voyages pour rien du tout*]^{N}
‘[And as the guys didn’t want to stay at sea] [(and) they wanted to go back to land every night] uh [I think we guzzled a lot of gas making the trips all for nothing]’ [source: FFAMNN01]

We can notice in this last example, that contrary to the rule in standard French, the conjunction *comme* 'as' is not picked up by *que* 'that' before the second verbal construction, which is linked to that which precedes by the non-final intonation of the latter.

b. Nucleus Post-Nucleus

The Nucleus Post-Nucleus configuration tends to be less common. Like the Pre-Nucleus, the Post-Nucleus fulfils various pragmatic functions depending on the microsyntactic content: antitopic in (31), discourse marker which addresses the interlocutor in (32), and specification of assertive force in (33) and (34):

(31) [*Ah oui non* Cn→ *mais c'est une folle* C0↓]^{N} [*elle* C0n←]^{S}
'[Ah yes no but she's a nutcase] [she is]' [source: FFAMCV11]

(32) [*Parce que en fait je croyais que c'était samedi*]^{N} [*le quatorze juillet*]^{S}
'[Because in fact I thought it was Saturday] [July 14]' [source: FFAMCV11]

(33) [*C'est Narcisse* C0↓]^{N} [*tu vois* C0n←]^{S}
'[It's Narcisse] [you know]' [source: FFAMCV11]

(34) [*Mais c'est une folle* C0↓]^{N} [*je t'assure* C0n←]^{S}
'[But she is crazy] [I assure you]' [source: FFAMCV11]

We note that there are rarely two Post-Nuclei and that while there are no constraints on their grammatical constitution, they are often built on a simple constituent, a pronoun, a noun, or a simple verbal construction.

3.3 Configurations as extended discourse patterns: Grouping

Nuclei can be grouped into larger configurations or discourse patterns by means of conjunctions in their use as discourse connectives, as in (35). This configuration, on the other hand, occurs frequently in all oral texts.

a. Nucleus – Discourse Connective – Nucleus

(35) [*D'ailleurs* C1↗]^{S} [*elle est allée à Saint-Trop* C2↘ *avec eux* C1↗]^{N} [*parce qu'ils ont des grosses motos* C1↗]^{N}
'[Besides] [she went to St. Trop with them] [because they have big bikes]' [source: FFAMCV11]

The grouping of two Nuclei can also be a consequence of textual organization. The utterances forming the main line of a discourse are integrated into larger discourse units such as narrative patterns, as in (36).

b. Nucleus Nucleus

(36) [*Et euh elle revient* C1↗] $^{\text{N}}$ [*deux minutes après* Cn→ *elle lui donne* Cn→ *son numéro de téléphone* C1↗ *et tout* C1↗]$^{\text{N}}$ [*rancart* C1↗ *pour mardi soir* Cn→ *et tout ça* Cn→ *quoi* C0↓]$^{\text{N}}$
‘[And she comes back] [two minutes later she gives him her phone number and everything] [date for Tuesday night and all that what’s the point]’
[source: FFAMCV11]

But Nuclei can also be grouped with another configuration in which the different Nuclei do not have the same function.

c. Nucleus [Nucleus] Nucleus

In (37) the second Nucleus is not part of the current discourse sequence and functions at another level that can be named the metadiscourse level in which the speaker verbalizes a pronunciation difficulty:

(37) [*On a fait une qui est le*]$^{\text{N1 part1}}$ [*c’est des noms anglais j’arrive pas bien à les prononcer ces machins-là*]$^{\text{P}}$ [*le Thistlegorm*]$^{\text{N1 part2}}$
‘[We did one that is] [they’re English names I can’t pronounce them well this stuff] [the Thistlegorm]’ [source: FFAMNN01]

In (37) the insertion lies between the two parts of the Nucleus, but there are also cases where a Nucleus is inserted between two Nuclei according to a foreground *versus* background relationship which is underlined by the tense concord. In (38), the narrative pattern is momentarily broken by the insertion of explanatory background information (in bold):[11]

(38) [*Elle est* Cn→ *elle est euh elle est partie* Cn→ *à la douche* C2↘ *deux minutes* C1↗] $^{\text{N1}}$ [***il y a il y a il y en a un*** Cn→ ***que je connaissais de vue*** C1↗***d’ailleurs*** C1↗] [***parce que j’avais*** C2↘ ***il avait bossé au Mac Do*** Cn→ ***avec moi*** C1↗] $^{\text{NPARENT}}$ // [*et euh elle revient* C1↗] $^{\text{N2}}$ [*deux minutes après* C1↗] $^{\text{S}}$ [*elle lui donne son numéro de téléphone et tout* C1↗] $^{\text{N3}}$ [*rancart* C1↗ *pour mardi soir et tout ça quoi* C0↓]$^{\text{N}}$
‘[She went to the shower for two minutes] [***there there there is somebody that I knew by sight also***] [***because I had he had worked at McDonalds with me***] [and she comes back] [two minutes later] [she gives him her phone number and everything] [date for Tuesday night and all that what’s the point]’
[source: FFAMCV11]

[11] See Debaisieux and Martin (2010) for a study of parenthetical utterances in spoken French.

In this example, the prosody groups several Nuclei into a paragraph, a “period”, to use the term of the Groupe de Fribourg (2003), whose last Nucleus is marked by a final intonation (C0↓). But the autonomy of the components also allows the prosody to “ungroup” a single macrosyntactic unit into several components. This is what we will analyze in the next section.

3.4 Ungrouping: Epexegesis

The term “epexegesis” was defined by Bally (1950, p. 57) as “the addition of a ‘monorem’ with a prepositional value intended to complete, to explain after the event the first utterance”.[12] Martins-Baltar (1977) defined epexegesis as “information added after the fact to an utterance in which it could have been syntactically integrated” (p. 23).[13] Here are two Examples (39)–(40) from the conversation:

(39) [*Ça fait un peu mal* C1↗] [*quand même* C1↗] [*qu'il te dise ça* C0↓]^{N} [***plutôt que*** C1↗] [***genre*** C2↘ ***ben tu me manques tout court*** C1↗] [*quoi* C0↓]^{N}
‘[It's a bit hurtful] [still] [he tells you that] [***rather than***] [***like well I miss you***] [what's the point]’ [source: FFAMCV11]

(40) *DEL: [*Donc on va se faire* Cn→ *une petite bouffe* Cn→ *dessus le soir* C1↗]^{N}
[*et on va regarder* Cn→ *le feu d'artifice* C0↓]^{N} <[qui] >
‘[So we'll have a bit of nosh in the evening] [and we will watch the fireworks] <[that]>’
*EST: < *Top* >
‘<Great>’
*DEL: [***qui est projeté du port*** C0↓]^{E}
[***that are let off in the harbor***]’ [source: FFAMCV11][14]

The structures in the epexegesis (in bold in the examples) constitute a specific configuration whose properties can appear contradictory. Prosodically, they clearly show enunciative autonomy: In (39), the construction is preceded by a major prosodic boundary, indicated in the example by the symbol C0↓. In (40), the construction is, moreover, enunciated in isolation, since it is preceded by a comment (‘great’) by the interlocutor. But while the construction is materially isolated from

[12] “L'adjonction d'un monorème à valeur prépositionnelle destiné à compléter, à expliquer après coup la première énonciation”. (Bally, 1950, p. 57)

[13] “Information ajoutée après coup à un énoncé auquel elle aurait pu s'intégrer syntaxiquement”. (Martins-Baltar, 1977, p. 23)

[14] In this example, the two participants are referred to by *DEL and *EST in accordance with the transcription conventions of C-ORAL-ROM.

the preceding construction, it is nevertheless syntactically dependent on the latter. There is no contradiction in the fact that a syntactic relation straddles a prosodic boundary when one considers, as we do, that the two components are autonomous.

In this description, the use of the two syntactic components, micro- and macrosyntax, associated with the prosodic component allows us to report on all the productions and to note some recurrent associations.

3.5 Some remarkable configurations in monologue

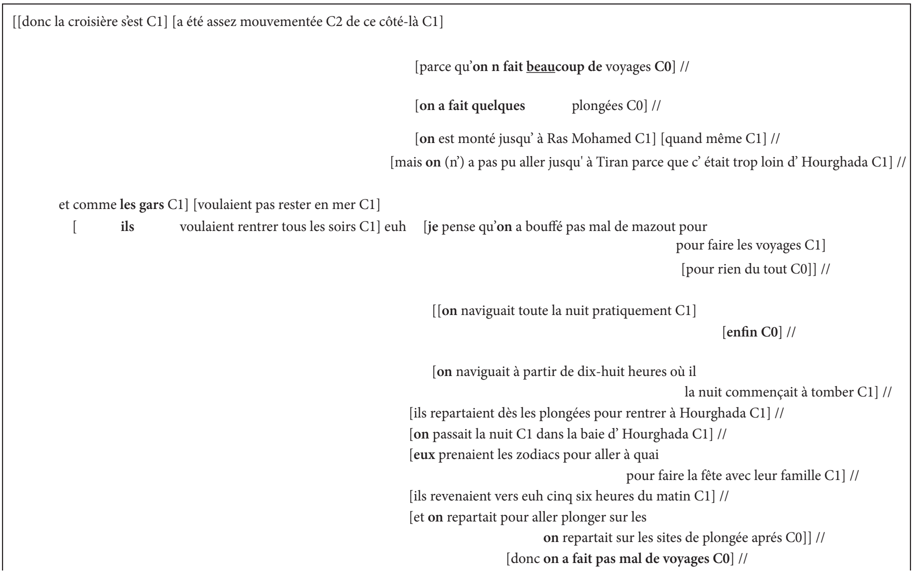
We end this paper by analyzing a monologue extract from a non-professional speaker: a travelogue about a cruise. We will see that it is very different, with respect to the relation between the three components microsyntax, macrosyntax and prosody, from speech by professional speakers, such as the politicians analyzed by Martin (2009). In this excerpt we have identified two parts we call "Cruise" and "Dive". They are presented separately here due to their length but the speaker built this part of his travelogue in a single block. For each part we present first the text and the English translation. However, in order to highlight the organization of the two passages of this monologue, we will present it as a "grid" (see Figures 1 and 2), developed by Blanche-Benveniste and Jeanjean (1987). This presentation allows the reader to see how the speech progresses both on the paradigmatic and the syntagmatic dimensions. This simplified presentation highlights the discursive organization and shows how syntax, prosody and lexicon contribute to structuring the sequence.

(41) a. Part 1: Cruise

Donc la croisière s'est a été assez mouvementée de ce côté-là parce qu'on a fait beaucoup de voyages // on a fait quelques plongées // on est monté jusqu'à Ras Mohamed quand même mais on (n') a pas pu aller jusqu'à Tiran parce que c'était trop loin d'Hourghada // et comme les gars voulaient pas rester en mer ils voulaient rentrer tous les soirs euh je pense qu'on a bouffé pas mal de mazout pour pour faire les voyages pour rien du tout // on naviguait toute la nuit pratiquement enfin // on on naviguait à partir de dix-huit heures où il la nuit commençait à tomber // ils ils re ils repartaient dès les plongées pour rentrer à Hourghada // on passait la la nuit dans la baie d'Hourghada // eux prenaient les zodiacs pour aller à quai pour faire la fête avec leur fa- leur famille // ils revenaient vers euh cinq six heures du matin // et on repartait pour aller plonger sur les on repartait sur les sites de plongée après // donc on a fait pas mal de voyages //

'So the cruise is was quite hectic from that point of view because we made many trips // we did some dives // we climbed up to Ras Mohamed anyway but we could not go up to Tiran because it was too far from Hurghada //

> and as the guys didn't want to stay at sea they wanted to go back to land every night uh I think we guzzled a lot of gas making the trips all for nothing // we sailed all night practically finally // we were sailing from six in the evening when night began to fall they // they left after the dives to go back to Hurghada // we spent the night in Hurghada Bay // they took the dinghies to go to the dock to party with their family // uh they returned about five or six o'clock in the morning and we set off again // to dive again on one of the dive sites after // so we did a lot of traveling but finally //'



[[donc la croisière s'est C1] [a été assez mouvementée C2 de ce côté-là C1]

[parce qu'**on n fait beaucoup de** voyages **C0**] //

[**on a fait quelques** plongées C0] //

[**on** est monté jusqu' à Ras Mohamed C1] [quand même C1] //

[mais **on** (n') a pas pu aller jusqu' à Tiran parce que c' était trop loin d' Hourghada C1] //

et comme **les gars** C1] [voulaient pas rester en mer C1]

[**ils** voulaient rentrer tous les soirs C1] euh [**je** pense qu'**on** a bouffé pas mal de mazout pour

pour faire les voyages C1]

[pour rien du tout C0]] //

[[**on** naviguait toute la nuit pratiquement C1]

[**enfin C0**] //

[**on** naviguait à partir de dix-huit heures où il

la nuit commençait à tomber C1] //

[ils repartaient dès les plongées pour rentrer à Hourghada C1] //

[**on** passait la nuit C1 dans la baie d' Hourghada C1] //

[**eux** prenaient les zodiacs pour aller à quai

pour faire la fête avec leur famille C1] //

[ils revenaient vers euh cinq six heures du matin C1] //

[et **on** repartait pour aller plonger sur les

on repartait sur les sites de plongée aprés C0]] //

[donc **on a fait pas mal de voyages C0**] //

Figure 1. Grid of "Cruise"

Syntax separates the different episodes. The beginning of each syntactic unit is materialized by a new line except if there is a list effect: The constituents displaying the same syntactic function are listed one under the other starting from the syntactic slot. This part of the story presents an explanatory sequence. It begins with a statement and its causal explanation built on a canonical complex construction which forms a Nucleus. But the reason for the first statement *on a fait pas mal de voyages* 'we made many trips' is developed in a long list of simple verbal constructions with a discursive topic that moves continuously forward by means of the subjects of these descriptive verbs *on/ eux/ ils* and alternating protagonists, introduced by the NP 'the guys' and emphasized by passing it to them.[15] The syntactic cohesion

[15] The subject *je* 'I' of the evaluative verb *penser* 'think' does not in fact interrupt the progression as it functions at another level of textuality.

of the whole episode is underscored by the constant use of the imperfect tense *on naviguait* 'we were sailing' and the recurrence of the verbal lexeme *repartait* 'we left'. The episode is constructed by means of a supporting sequence with a looping effect that can be schematized as follows:

a. Opening line: cruise hectic because many trips
b. Development: nighttime adventures
c. Conclusion of the demonstration: so we made many trips.

Semantically, the organization of the passage can be compared to an increase of information achieved by constant theme progression (see Combettes, 1993) as shown by the same subjects *on* 'we' with different action verbs. Note that most of the Nuclei consist of simple verbal constructions. There is only one case where the Nucleus is preceded by two Satellites which have an explanatory value. In general, there is most often congruence between micro- and macrosyntax. On the other hand, we note that the relation between prosody and macrosyntactic units is of the enclosing type. Several Nuclei, in particular the list of Nuclei with a descriptive value, are linked by a continuous intonation, and only the conclusion is marked by a final intonation. This suggests that the systematic repetition of the lexicon and constructions makes it possible to compensate for the listener's difficulty in retaining such extended prosodic groups.

This structure is quite different from the second part of the transcript which presents a descriptive sequence.

(41) b. Part 2: Dives
Mais enfin donc les p par contre les les plongées qu'on a fait étaient superbes ait pas mal d'épaves on a été faire donc des un site où il y a les les épaves // on a fait entre autres le Giannis Day le Carnatic // on a fait une qui est le (c'est des noms anglais j'arrive pas bien à les prononcer ces machins-là…) le Thistlegorm // c'est un bateau euh de de guerre qui a été coulé donc par les les Allemands en dix-neuf cent quarante // et il est sympa à faire parce que c'était un transport de matériel // c'est un bateau qui est sur quarante mètres de fond environ // enfin il est il la coque le le sable est à quarante mètres // le le haut du bateau est plus haut il est sur trente mètres // et dans les cales tout est encore ar arrimé xxx // on voyait vraiment il y avait tous les les camions les les jeeps tout ça // c'est tout aligné dans les cales // c'est tout rangé les unes dé- à côté des autres bien bien alignées // et dans les camions il y a à l'intérieur des camions bon ils ils ils optimisaient toutes les les places donc de sur sur ce type de transport // dans les bennes de camions il y avait toutes les motos qui étaient entassées et toutes alignées comme ça // il y a des centaines de motos sur cet(te) sur ce bateau alignées dans les dans les cales de dans les bennes de camions //

> ‘On the other hand the dives that we did were great lots of wrecks so we did a site where there are the wrecks // one of them we did was the Carnatic the Giannis Day // we did one that is (they’re English names I can’t pronounce them well this stuff then…) the Thistlegorm // it is a uh warship that was sunk by the Germans in nineteen forty // and it was fun to do because it was a cargo vessel // it’s a boat that is sitting about forty meters deep // well it is the hull the sand is forty meters deep // the top of the boat is above it is at thirty meters // and in the holds everything is still secured xxx // we really saw there were all the trucks jeeps all that // it’s all aligned in the holds // it’s all stored one next to the other well aligned // and trucks there inside the trucks well they they they optimized all the the space on this type of transport // in the dumpers there were all the motorcycles that were stacked and aligned all like that // there are hundreds of motorbikes on this on this boat lined up in the hold in the dumper trucks //’

[mais enfin donc **par contre** **les plougées** qu’on a fait **étaient superbes** C0] //
[**on a fait pas mal d’ épaves C1**] //
[**on a été faire** donc un site **où il y a les épaves** C1] //
[**on a fait** entre autres le Giannis Day C1]
[le Camatic C1]
[**on a fait** une qui est le
[***c’est des noms anglais j’arrive pas bien à les prononcer ces machins-là C0***]
[**le Thistlegorn** C0]] /
c’est **un bateau** euh de guerre C1] [qui a été coulé donc par les Allemands Cn en 1940 C1] //
et [il est sympa à faire C1] [parce que c’était un transport de matériel C0]] //
[**c’est un bateau qui** est sur quarante mètres de fond C1] [environ C1] //
[enfin **la coque** le sable C2 **est à quarante mètres** C1] //
[**le haut du bateau** est plus haut C1] //
[il est **sur trente mètres** C1] //
[et **dans les cales** C1] / **tout** est encore **arrimé C0**] //

[on voyait vraiment C1]
[il y avait **tous les camions** C1]
[**les jeeps** C1]
[tout ça C0] //
[**c’est tout aligné** dans les cales C1] //
[**c’est tout rangé** C1] [les unes à côté des autres C1] [
bien aligné(es) C1] //
[**et dans** **les camions** C1] /
[**à l’intérieur des camions** Cn] /
[***bon ils optimisaient toutes les places C1 sur ce type de transport C1***]
[**dans** **les bennes de camions** C1] / [il y avait **toutes les motos** C1] [qui étaient **entassées** C1]
et toutes **alignées** C1] [comme ça C1]//
[ily a **des centaines de inotos** C1] [sur ce bateau C1] /
[**alignées** C1] [**dans les bennes de camions C0**] //

Figure 2. Grid of “Dives”

The second part of the passage concerns dives. A very brief narrative sequence marked by the same principle of continuous theme (see Combettes, 1993) and repetitions of the verbal construction *on a fait* ‘we did’ results in naming the object of a dive “the Thistlegorm”. After a brief evaluative commentary (again final accent) the speaker builds up a description by a progressive zoom effect achieved by means of “expanded” theme progression: the boat (hull, top of the boat), then the holds of the boat, then the trucks in the holds. Each theme is presented by a Satellite with framing value which introduces each subsequence: the boat, the holds of the boat, and the trucks in the holds. The object of the description is globally located 'everything is still secured'. The final accent closes on an end-of-list element *tout ça* 'all that'. The speaker then continues with a subsequence semantically organized according to a linear theme: the "rhematic" introduction of the object to be described by a presentational structure, 'there were all the trucks', then taken up by another Satellite built on a detached element, 'in the trucks', which introduces a new object rhetorically presented by the presentative 'there were all the motorcycles'. To these effects of "anadiplosis" or "lexical repetitions and sliding to the left"[16] is added a chiasmus that concludes the sequence: *dans les bennes de camions il y avait toutes les motos qui étaient entassées et toutes alignées comme ça // il y a des centaines de motos […] alignées […] dans les bennes de camions* 'in the dumpers there were all the motorcycles that were stacked and aligned all like that // there are hundreds of motorbikes on this on this boat lined up in the hold in the dumper trucks'. In addition, there are many lexical and grammatical repetitions. Here too, various Nuclei are grouped in a long paragraph and only the final Nucleus has a final prosody. Each final Nucleus has a particular value. The first line expresses a general opinion on the dives. The second "final Nucleus" closes the enumeration of the various wrecks and creates an expectation by specifying the name of the last wreck mentioned. The third "final Nucleus" closes a brief presentation of this wreck. Then the final Nuclei coincide almost with each place introduced by the Satellites with a framing value.

These sequences are very similar to what Moneglia and Raso (2014) call *Stanza*,[17] that is, sequences that are

> characterized by weak assertive illocutionary forces… and contain prosodic cues (tails) indicating that the discourse goes on until the sequence reaches a conclusion which is signaled by a terminal prosodic break. (p. 489)

What we think is important to point out is the role played by the repetitions of structures and lexis as anchoring devices for the discourse, thus maintaining the state of the "discursive memories" (Berrendonner, 2003) despite the absence of strong prosodic boundaries and allowing the listener to process sequences of a reasonable size.

By contrasting two types of styles, monologue and conversation, we see that while the units involved are of the same nature, the discourse patterns vary in shape.

[16] Title of the article by C. Blanche-Benveniste (1993) in which she conducts a syntactic analysis of the aspectual effects of progression of certain repetitions.

[17] The term is due to Cresti & Moneglia (2010).

The monologue displays basic combinations of macrosyntactic units built around a syntactic Nucleus matching with a terminal prosodic contour that defines the illocutionary force of the utterance. The few Satellites have a discursive value: They give an explanation for a statement or build a frame in a descriptive organization. We have only one non-verbal Satellite with a topic value in the first description of the wreck: *la coque le le sable est à quarante mètres* 'the hull the sand is forty meters deep'. These stanzas encompass many elementary macrosyntactic units and are structured by a whole range of "folk" rhetoric devices: repetitions of syntactic structures and lexical units as anchoring devices for the discourse, rhetorical tropes such as chiasmus. These formal devices help maintain a basic level of coherence in the discourse, despite the absence of strong prosodic boundaries and therefore allow the listener to process pieces of information of reasonable size.

In the conversation, while we also find a certain number of Nuclei made up of canonical verbal constructions, we generally find more configurations as Satellite Nuclei. But the difference is also in the function of Satellites. Post-Nuclei Satellites seem to be used to handle interaction since they express assertive force or address the interlocutor. Pre-Nuclei are used as framing devices but also to change topic. Epexegesis allows the speaker to continue after an interruption. We note also that prosody functions more than in the monolog to "ungroup" single macrosyntactic units into several components, and that there are more non-verbal Nuclei with an emphatic value.

4. Conclusion

In the theoretical part of our paper we set out a framework defining two combinatorial principles, macrosyntax and prosodic structure, on the basis of which speakers shape their utterances. These two combinatorial devices do not necessarily match, which considerably enlarges the resources available to speakers. However, the macrosyntactic component allows us to characterize elementary units, which can be considered as the building blocks of the formal patterns found in discourse.

In the descriptive part, the comparison between the units identified in the conversation and those found in the monologue shows the fundamental role of prosody and Satellites in moving the discourse forward in the monologue and in managing the interaction in conversation. A more systematic analysis is of course necessary to confirm these findings.

Acknowledgements

We are grateful to Shlomo Izre'el and to anonymous reviewers for their useful comments on an earlier version of this work.

References

Bally, C. (1950). *Linguistique française et linguistique générale*. Bern: A. Francke.

Berrendonner, A. (2003). Éléments pour une macro-syntaxe: Actions communicatives, types de clauses, structures périodiques. In A. Scarano (Ed.), *Macro-syntaxe et pragmatique: l'Analyse linguistique de l'oral* (pp. 93–110). Roma: Bulzoni.

Blanche-Benveniste, C. (1993). Répétitions de lexique et glissement vers la gauche. *Recherches sur le Français Parlé*, 12, 9–34.

Blanche-Benveniste, C. (2010). *Approches de la langue parlée en Français*. Paris: Ophrys.

Blanche-Benveniste, C., Bilger, M., Rouget, C., Van den Eynde, K., Mertens, P. (1990). *Le français parlé: Études grammaticales*. Paris: Editions du CNRS.

Blanche-Benveniste, C., & Jeanjean, C. (1987). *Le français parlé: Transcription et édition*. Paris: Didier Érudition.

Blanche-Benveniste, C. & Martin, P. (2010). *Le français. Usages de la langue parlée*. Leuven/Paris: Peeters.

Combettes, B. (1993). *Pour une grammaire textuelle*. Louvain: De Boeck-Duculot.

Cresti, E., & Moneglia, M. (2010). Informational patterning theory and the corpus based description of spoken language. In M. Moneglia & A. Panunzi (Eds.), *Bootstrapping information from corpora in a cross linguistic perspective* (pp. 13–45). Florence: Firenze University Press.

Culicover, P., & Jackendoff, R. (2005). *Simpler syntax*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199271092.001.0001

Culioli, A. (1983). Pourquoi le français parlé est-il si peu étudié?. *Recherches sur le Français Parlé*, 5, 291–300.

Debaisieux, J.-M, (2016). Toward a global approach to discourse uses of conjunctions in spoken French. *In Language Sciences 58*. Special Issue. Adverbial Patterns in Interaction. Dagmar Barth-Weingarten and Oliver Ehmer (ed), 79–94. https://doi.org/10.1016/j.langsci.2016.04.001

Debaisieux, J-M. (Ed.). (2013). *Analyses linguistiques sur corpus: Subordination et insubordination en français*. Paris: Hermès-Lavoisier.

Debaisieux, J-M., & Deulofeu, J. (2001). Grammatically unacceptable utterances are communicatively accepted by native speakers, why are they? *Disfluency in Spontaneous Speech (DiSS '01)*, 69–72.

Debaisieux, J-M., & Martin, P. (2010). Les parenthèses: Étude macrosyntaxique et prosodique sur corpus. In M-J. Béguelin, M. Avanzi, & G. Corminboeuf (Eds.), *La parataxe: Vol. 2. Structures, marquages et exploitation discursive* (pp. 307–339). Bern: Peter Lang.

Debaisieux, J-M., Deulofeu, J., & Martin, P. (2019). Apparent insubordination as discourse patterns in French. In K. Beijering, G. Kaltenböck, & M. S. Sansiñena (Eds.), *Insubordination: New perspectives* (pp. 234–257). Berlin: De Gruyter Mouton.

Deulofeu, J. (2003). L'approche macrosyntaxique en syntaxe: Un nouveau modèle de rasoir d'Occam contre les notions inutiles. *Scolia*, 16, 47–62.

Deulofeu, J. (2008). Peripheral constituents as generalized hanging topics. In R. Kawajima, G. Philippe, & T. Sowley (Eds.), *Phantom sentences: Essays in linguistics and literature presented to Ann Banfield* (pp. 227–257). Bern: Peter Lang.

Deulofeu, J. (2013). L'approche macrosyntaxique: Sources et controversies. In J.-M. Debaisieux (Ed.), *Analyses linguistique sur corpus: Subordination et insubordination en Français* (pp. 427–498). Paris: Hermès-Lavoisier.

Groupe de Fribourg. (2013). *Grammaire de la période*. Bern: Peter Lang.

Halliday, M. (1989). *Spoken and written language*. Oxford: Oxford University Press.

Martin, P. (2009). *L'intonation du français*. Paris: Armand Colin.

Martin, P. (2014). Spontaneous speech corpus data validates prosodic constraints. In N. Campbell, D. Gibbon, & D. Hirst (Eds.), *Proceedings of the 6th conference on speech prosody* (pp. 525–529). Dublin: Science Foundation Ireland.

Martin, P. (2015). *The structure of spoken language: Intonation in romance*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139566391

Martin, P. (2018). *Intonation, structure prosodique et ondes cérébrales*. London: ISTE.

Martins-Baltar, M. (1977). *De l'énoncé à l'énonciation: Une approche des fonctions intonatives*. Paris: CREDIF.

Mithun, M. (2005). On the assumption of the sentences as the basic unit of syntactic structures. In Z. Frayzingier, A. Hodges, & D. S. Rood (Eds.), *Linguistic diversity and language theory* (pp. 169–183). Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.72.09mit

Mithun, M. (2008). The extension of dependency beyond the sentence. *Language*, 84(1), 69–119. https://doi.org/10.1353/lan.2008.0054

Moneglia, M., & Raso, T. (2014). Notes on Language into Act Theory (L-AcT). In T. Raso & H. Mello (Eds.), *Spoken Corpora and Linguistic Studies* (pp. 468–495). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61.15mon

Partee, B. (1995). Lexical semantics and compositionality. In L. R. Gleitman, M. Liberman, & D. N. Osherson (Eds.), *Language: Vol. 1. An invitation to cognitive science* (pp. 311–360). Cambridge, MA: The MIT Press.

Rossi, M. (1971). Le seuil de glissando ou seuil de perception des variations tonales pour la parole. *Phonetica*, 23, 1–33. https://doi.org/10.1159/000259328

Verstraete, J-C. (2007). *Rethinking the coordinate-subordinate dichotomy: Interpersonal grammar and the analysis of adverbial clauses in English*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110918199

CHAPTER 5

Design and annotation of two-level utterance units in Japanese

Takehiko Maruyama[i,iii], Yasuharu Den[ii,iii], and Hanae Koiso[iii]
[i]Senshu University / [ii]Chiba University / [iii]National Institute for Japanese Language and Linguistics

We introduce an annotation scheme of two-level utterance units in Japanese speech, thus identifying utterance units in two different levels, which are called "short utterance-unit" (SUU) and "long utterance-unit" (LUU). SUUs are divided by acoustic and prosodic boundaries, corresponding to Intonation Units (Chafe, 1994), considered as basic units of speakers' planning. LUUs, on the other hand, correspond to Clausal Units (Biber, Johansson, Leech, Conrad, & Finegan, 1999), being divided by major syntactic breaks and/or communicative interactions. Those are basic units of syntactic chunks and/or participants' interaction. We show a design of SUU and LUU consisting of prosodic, clausal and non-clausal units. Annotating SUU and LUU in 12 dialogs of two hours altogether, we examine their characteristics and distribution in the corpus.

Keywords: two-level utterance-units, short utterance-unit (SUU), long utterance-unit (LUU), acoustic and prosodic boundaries, syntactic breaks, communicative interactions, speaker's planning, participants' interaction

1. Introduction

For linguists who study spoken discourse the question of how to divide the flow of speech to extract standard, useful, and effective units in a uniform manner has always been a problem. In written text, sentence-final boundaries are usually marked by periods, and these identify the extent of each sentence reliably. In spoken discourse, on the other hand, there are no explicit devices to mark utterance-final boundaries, and this often causes difficulties for defining basic units of speech. Dialogs often consist of short and fragmentary messages, while monologs and narratives in dialogs sometimes form very long and complicated structures. Nonetheless, we have an intuition that there are some fundamental units of speech that form basic segments at a certain level higher than words and

https://doi.org/10.1075/scl.94.05mar
Additional files available at http://doi.org/10.1075/scl.94.05mar/audio

phrases. If we call them “utterances”, then what type of boundary can be treated as a reasonable cue to identify such basic speech units, and how can they be defined? Should we expect to find linguistically unique and homogeneous units, or heterogeneous ones with various levels of linguistic structures? The development of an established scheme for annotating such units is a crucial step towards corpus-based studies of spoken discourse and dialog.

Several attempts have been made to define utterance units from various aspects, including prosody (Beckman & Ayers, 1994; Du Bois, Shuetze-Coburn, Cumming, & Paolino, 1993; Iwasaki, 1993; Venditti, 1994), syntax (Meteer et al., 1995), and pragmatics (The AMI Project, 2005). Yet, we still do not have a widely-used scheme for identifying fundamental units in dialogs. Ford and Thompson (1996) analyzed the interrelationship among prosodic, syntactic, and pragmatic/action completion points of utterances in English conversations, showing that the majority of speaker changes occurred at *complex transition-relevance places*, which are defined by the convergence of prosodic, syntactic, and pragmatic/action completions. This result suggests that utterance units suitable for various dialog research should be defined in a complex way by taking prosody, syntax, and pragmatics into account simultaneously.

In this paper, we propose a new scheme for annotating utterance-units in Japanese speech, especially dialogs. In Section 2, we first apply four separate criteria, partitioning the flow of speech into (1) inter-pausal units, (2) intonation units, (3) clause units, and (4) pragmatic units, respectively. We annotate Japanese dialog data across these four criteria and then analyze the interrelationships among the four unit types by using correspondence analysis and cluster analysis. In this way, we show that the distributions of the annotated labels of these units can be classified into several groups according to the depth of unit boundary. Based on these results, in Section 3, we come up with an annotation scheme that integrates the four unit types, distinguishing two sorts of utterance-units with different granularities: *short and long utterance-units.* Short utterance-units are identified by pauses and intonation breaks, whereas long utterance-units are identified by syntactic and pragmatic boundary. Applying this scheme to our dialog data, we explore some characteristics of these utterance-units in Section 4, focusing particularly on unit duration and syntactic property, as well as on hearers’ responses. Finally, in Section 5, we introduce an extension of our scheme, considering interactions between the speaker and the hearer, and mismatches between short and long utterance units. We conclude the paper by discussing how our two-level utterance-units are useful in analyzing cognitive and communicative aspects of spoken dialogs.

2. Analysis of the interrelationships among four utterance-unit types

2.1 Data

In this study, we used two dialog corpora, the *Chiba Three-Party Conversation Corpus* (Den & Enomoto, 2007) and the *Corpus of Spontaneous Japanese* (Maekawa, 2003). The *Chiba Three-Party Conversation Corpus* (henceforth, Chiba corpus) is a collection of 12 casual conversations recorded at Chiba University, each conducted by three campus friends. The *Corpus of Spontaneous Japanese* (henceforth, CSJ), on the other hand, is a huge speech corpus of 651 hours with 7.52 million words, containing a large amount of monologs as well as a small amount of dialogs. The dialog part of the CSJ consists of dyadic conversations between interviewers and interviewees and contains 12.2 hours of recorded speech and 150,000 words.

Four dialogs from the Chiba corpus and another four dialogs from the CSJ were used for the current study. The former consists of a total of 38 minutes with 9,099 words, and the latter consists of 48.3 minutes with 10,558 words. For each dialog, a five-minute fragment, beginning one minute after the start of the dialog, was extracted for annotation and analysis. Thus, eight dialog fragments totaling 40 minutes in length with 9,235 words were used in the current study. All dialogs were carefully and precisely transcribed (including fillers, word fragments, various disfluencies, laughter, and coughing), and manually segmented into words with time information supplied at every word boundary.

2.2 Annotation

Three preexisting utterance unit types, (1) inter-pausal units (IPUs), (2) intonation units (IUs), and (3) clause units (CUs), as well as a newly created one, (4) pragmatic units (PUs), were identified and their boundary labels were annotated. With the exception of IPUs, for which annotation was automatic, the annotation for each unit type was performed by a different non-expert annotator, and crosschecked by at least one of the authors. Table 1 summarizes the annotation labels of these units.

2.2.1 *Inter-pausal units (IPUs)*

Inter-pausal units (IPUs) (Koiso, Horiuchi, Tutiya, Ichikawa, & Den, 1998) were automatically identified by making reference to the time-stamps in the word-segmented transcripts. A stretch of speech followed by a pause longer than 100 ms was recognized as an IPU.

2.2.2 *Intonation units (IUs)*

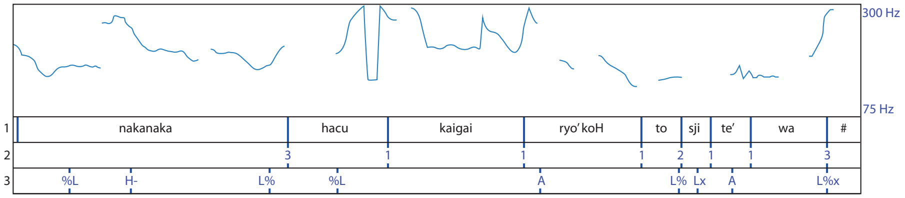
Intonation units (IUs) were labeled based on the X-JToBI scheme (Maekawa, Kikuchi, Igarashi, & Venditti, 2002), an extension of the standard JToBI (Venditti, 1994) for spontaneous speech. By judging perceived intonational boundary, a break index (BI) was assigned to every word boundary, with strengths of BI indicated by numbers, 1, 2, or 3. When a boundary with BI = 2 was followed by a perceived pause, BI = 2p was used instead. A stretch of speech delimited by boundaries with BIs greater than or equal to 2 was recognized as an IU, which roughly corresponds to an accentual phrase. A final boundary tone, L%, H% (LH%), or HL% (LHL%), was also associated with each IU. Fillers (along with certain types of interjection) and disfluencies were labeled with special marks, "F" and "D", respectively. Figure 1 shows an example of annotated IU labels. Words, break indexes, and final boundary tones are annotated in the tiers 1, 2, and 3, respectively.

Table 1. Annotation labels

Inter-Pausal Unit (IPU)	
100	Followed by a pause longer than 100 msec
Intonation Unit (IU)	
3-H%	BI = 3, Tone = H%, LH%
3-HL%	BI = 3, Tone = HL%, LHL%
3-L%	BI = 3, Tone = L%
2p-H%	BI = 2 + pause, Tone = H%, LH%
2p-HL%	BI = 2 + pause, Tone = HL%, LHL%
2p-L%	BI = 2 + pause, Tone = L%
2-H%	BI = 2, Tone = H%, LH%
2-HL%	BI = 2, Tone = HL%, LHL%
2-L%	BI = 2, Tone = L%
F	BI = F
D	BI = D
Clause Unit (CU)	
AB	Absolute boundary
SB	Strong boundary
WB	Weak boundary constituting a CU boundary
NB	Non-predicative boundary
MB	Unit-initial/final interjection
FB	Unit-initial/final word fragment

Table 1. (*continued*)

Pragmatic Unit (PU)	
c	Communicative modality
e	Epistemic/deontic modality
n	Null modality
f	Unit-initial fragment
B	Backchanneling response token
E	Expressive response token
L	Lexical response token
O	Response token of other type (repetition, completion, or assessment)
Br	Reply/acknowledgment with B form
Er	Reply/acknowledgment with E form
Lr	Reply/acknowledgment with L form
Or	Reply/acknowledgment with O form



nakanaka hatu kaigai ryokoo to-si-te-wa
fair first abroad travel as-TOP
fairly much, for a first trip abroad,

Figure 1. Annotation of IU boundaries.

2.2.3 *Clause units (CUs)*

Clause units (CUs) were originally designed to achieve segmentation of monologs (Takanashi, Maruyama, Uchimoto, & Isahara, 2003), and have been extended to cover dialog data (Maruyama, Takanashi, & Yoshida, 2010). Japanese is an SOV language, and the final boundary of a sentence is grammatically marked by a predicate, possibly followed by one or more auxiliary verbs and sentence-final particles. In colloquial Japanese, however, extremely long clausal chains are sometimes formed by concatenation of clauses using a variety of clause linkage markers, which results in a very long "sentence" without being accompanied by an explicit sentence final marker (Iwasaki & Ono, 2002; Maruyama, Horn, Russell, & Frellesvig, 2017). Thus, some sort of morpho-syntactic criteria should be adopted to segment the flow of speech into more tractable syntactic units.

The annotation scheme of CU identified four types of boundaries: absolute, strong, weak, and non-predicative boundaries (AB, SB, WB, and NB). ABs are characterized by explicit sentence final markers. SBs are characterized by conjunctive particles expressing coordination. WBs are characterized by other conjunctive particles followed by a discourse marker or speaker change. NBs are characterized by a turn's completion with no predicate. Two additional types of boundaries for unit-initial/final interjections and word fragments were also used: miscellaneous and fragmental boundaries (MB and FB). Table 2 shows an example of annotated CU labels.

In Table 2, "[" indicates the beginning point of an overlapping between two speakers' utterances, and "=" a latching between consecutive utterances.

2.2.4 *Pragmatic units (PUs)*

In addition to the three types of units described by acoustic (IPU), prosodic (IU), and morpho-syntactic (CU) features, another kind of unit was identified based on pragmatic properties. A pragmatic unit (PU) was defined as a unit that constitutes a single proposition, which was identified by modality expressions, response tokens, and short expressions functioning as replies/acknowledgments.

Linguistic modality in speech (defined as the speaker's mental attitude toward the propositional content and toward the hearer) was utilized to classify PUs. Three classes of linguistic modalities were distinguished: communicative (c), epistemic/deontic (e), and null (n) modalities. Communicative modality included not only explicit grammatical devices to signal speaker's attitude to the hearer such as sentence-final particles *ne* (seeking confirmation), *yo* (imparting new information), and *ka* (questioning), but also those expressed by rising intonation and implied by the context. Epistemic modalities were identified by modal auxiliary verbs like *daroo* 'probably', *hazu-da* 'is supposed to be', and *no-da* 'is the relevant information', and deontic modalities were identified by modal auxiliary verbs like *beki-da* 'ought to' and *nai-to-ike-nai* 'have to'. When there were no such modality expressions, null modality was annotated.

An additional notation was introduced to deal with fragments (f) due to false starts and self-interrupted speech. Furthermore, response tokens (Clancy, Thompson, Suzuki, & Tao, 1996; Den, Yoshida, Takanashi, & Koiso, 2011; Gardner, 2001), which are produced by a hearer during a speaker's turn, were recognized separately, and classified into four types: backchanneling interjections (B) such as *un* 'yeah' and *hai* 'yes', expressive interjections (E) such as *a* 'ah' and *hee* 'aha', lexical response tokens (L) such as *soo* 'I see' and *naruhodo* 'I understand', and other kinds of tokens (O), including repetitions of (part of) others' speech, collaborative completions, and assessments. When tokens with the same forms as these response tokens were used as replies to questions, requests, etc., or acknowledgments, they were labeled "Br", "Er", "Lr", and "Or". Table 3 shows an example of annotated PU labels.

Table 2. Annotation of CU boundaries.

Start	End	Sp	Transcript	Label
201.65	205.30	R:	nihon-ni kuru mae-mo mo[o nihongo perapera-da-si:= Japan-to come before-also already Japanese fluent-CP-and	SB
203.70	204.11	L:	[u:n yeah	MB
204.95	205.66	L:	=a: ah	MB
205.66	206.33	L:	soo-[na-n-[da I.see-COP-N-COP	AB
205.92	206.15	R:	[un yeah	MB
206.15	208.94	R:	[nihon-ni-mo ki-te: nihongo gakkoo-ni-mo Japan-to-also come-CP Japanese school-to-also it-te-[tte kanji-de go-CP-QP something.like.that-and	WB
208.28	209.19	L:	[un un un [un yeah yeah yeah yeah	MB
208.99	209.31	R:	[sono that	MB
209.35	209.45	R:	(D n) n-	FB
209.78	213.32	R:	kotoba-ni kansi-te-[wa anmari toraburu-ga language-DAT about-TOP so.much trouble-NOM maa nakat-ta-n-de like be.not-PAST-because	WB
210.77	211.16	L:	[u:n yeah	MB

Table 3. Annotation of PU boundaries.

Start	End	Sp	Transcript	Label
334.64	334.79	R:	so yeah	L
334.79	336.90	R:	dairen-kara pekin-ni-mo Dalian-from Beijing-to-also it-ta-n-desu-[kedo go-PAST-N-POL-but	e
336.63	337.07	L:	[ee e[e yeah yeah	B
336.90	338.57	R:	[sono toki-wa-ne densya-de that time-TOP-FP train-by zutto all.the.way	n
338.70	338.82	L:	a ah	E

(*continued*)

Table 3. (*continued*)

Start	End	Sp	Transcript	Label
339.02	339.64	L:	soo-na-n-[da I.see-COP-N-COP	L
339.44	339.72	R:	[un yeah	Br
339.85	340.55	L:	ta [ihen-desi-ta tough-POL-PAST	n
340.01	340.05	R:	[(D te) te-	f

2.3 Statistical analysis

The annotations of the four unit types above were combined into a tabular form, in which the four types of annotation labels were aligned at every word. When a unit had no boundary at a given word, a special label "*", indicating "no boundary", was assigned. A total of 9,266 words, 5,001 from the Chiba corpus dialogs and 4,265 from the CSJ dialogs, were obtained and used in the subsequent statistical analysis. In order to investigate the interrelationship among the four unit types, first, multiple correspondence analysis was applied to the aligned annotation labels. Multiple correspondence analysis produces a geographic configuration of labels from multiple factors, in which labels with similar distributions are located close together. Then hierarchical cluster analysis was conducted to classify annotation labels based on the three-dimensional coordinates obtained by the multiple correspondence analysis. The distance measure was Euclidean, and the agglomeration method was the Ward method.

2.4 Results

Figure 2 shows the cluster dendrogram for the annotation labels. The dendrogram shows the hierarchical relationship between annotation labels, indicating the order in which the clusters are joined. The heights reflect the distance between the clusters. We can obtain five major clusters, as shown by dashed rectangles in the figure, such that any two labels within any cluster are similar to each other but any two labels across clusters are dissimilar to each other. These five clusters can be characterized by the following features:

1. Syntactic and pragmatic disjuncture (CU = AB,SB,WB,NB; PU = c,e,L,O,Lr);
2. Acoustic and prosodic disjuncture (IPU = 100; IU = 3-,2p-);
3. Non-utterance boundary (IPU = *; IU = 2-,*; CU = *; PU = *);
4. Fragments (IU = D; CU = FB; PU = f);
5. Backchanneling and expressive interjections (IU = F; CU = MB; PU = B,E,Br,Er).

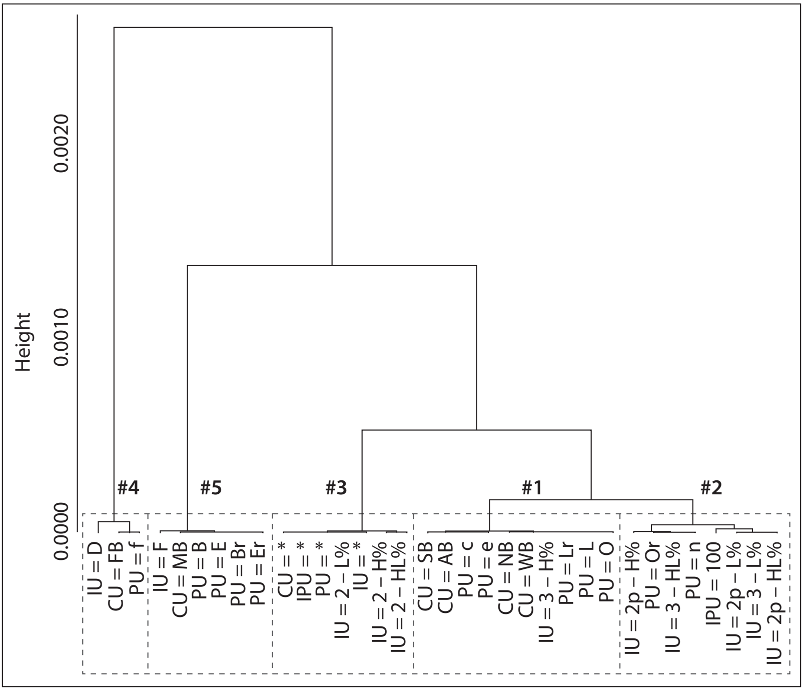


Figure 2. Cluster dendrogram for annotation labels

Cluster #1 can be characterized by syntactic boundary (CU = AB,SB,WB,NB) as well as pragmatic boundary (PU = c,e,L,O,Lr). Cluster #2, on the other hand, can be characterized by acoustic boundary (IPU = 100) as well as by prosodic boundary (IU = 3-,2p-). The labels included in clusters #4 and #5 are related to fragments and backchanneling/expressive interjections, respectively. The remaining cluster, #3, can be regarded as a "non-utterance boundary" cluster if we consider "IU = 2-" (i.e., accentual phrase boundary without boundary pitch movement or following pause) to not constitute an utterance boundary.

Three of these clusters, #1, #2, and #3, appear to be in order of the amount of disjuncture. Syntactic and pragmatic disjuncture is greater than acoustic and prosodic disjuncture, which is greater than no boundary. They can be put on a cline according to the amount of disjuncture. The remaining clusters, #4 and #5, seem better treated aside from this line. These results led us to an utterance-unit annotation scheme that integrates the four schemes we have discussed so far.

3. Proposal of a two-level annotation scheme

Now, we are in a position to propose our empirically emerged two-level annotation scheme for utterance-units in Japanese dialogs. Syntactic and pragmatic disjuncture are deeper unit boundaries, which consist of big structural breaks, whereas acoustic and prosodic disjuncture are shallower boundaries. It may be assumed that there is a hierarchical relationship between utterance-units determined by acoustics and prosody and those determined by syntax and pragmatics, the former being subsumed under the latter. Tentatively, we accept this assumption to obtain our two-level annotation scheme, although this assumption will be revisited in Section 5.2.

We refer to utterance-units defined by acoustic and prosodic disjuncture as *short utterance-units* (SUUs), and those defined by syntactic and pragmatic disjuncture as *long utterance-units* (LUUs). In addition, backchanneling and expressive interjections and fragments are identified separately, while being operationally included in both SUUs and LUUs. The procedures shown in Figure 3 enable us to recognize these utterance-units in dialogs. Table 4 sets out an example of our utterance-units, together with the underlying annotations of the four units. In Table 4, each row corresponds to a short utterance-unit. Long utterance-units can be obtained by concatenating rows labeled "S" with the succeeding rows.

[Boundary classification rules]

Apply the following rules in this order at every word boundary:

1. If the tokens so far constitute a fragment, mark the boundary with "F";
2. Else if the tokens so far constitute a backchanneling or expressive interjection, including those which function as a reply or an acknowledgment, mark the boundary with "R";
3. Else if the current boundary is a syntactic and/or pragmatic disjuncture, which may be expressed by a clause-unit boundary or a linguistic modality, mark the boundary with "L";
4. Else if the current boundary is an acoustic and/or prosodic disjuncture, which may be expressed by a pause or an intonation break, mark the boundary with "S";
5. Otherwise, apply these rules at the next word boundary.

[Unit identification rules]

Short utterance-units: Identify all four types of boundaries above as boundaries of short utterance-units.
Long utterance-units: Identify all types of boundaries above but "S" as boundaries of long utterance-units.

Figure 3. Rules identifying short and long utterance-units

Table 4. Example of the annotation of the four units (IPUs, IUs, CUs, and PUs) as well as the proposed utterance-units (UUs).

Start	End	Sp	Transcript	IPU	IU	CU	PU	UU
120.08	120.71	R:	kekkoo-ne fairly-FP	100	3-H%	NB	c	L
120.85	121.16	L:	nne FP	100	3-H%	NB	Lr	L
121.16	121.42	R:	un yeah	*	F	MB	Br	R
121.42	122.25	R:	na[kanaka rather	*	3-L%	*	*	S
121.66	121.88	L:	[(D in) im-	100	D	FB	f	F
122.29	123.89	R:	[hatu-kaigai-ryokoo-to-si-[te-wa first-overseas-travel-as-TOP	100	3-L%	NB	n	L
122.39	123.13	L:	[inpakuto-ga impact-NOM	100	2p-H%	NB	n	L
123.57	124.77	L:	[<laugh>	–	–	–	–	–
124.86	125.52	L:	naruhodo I.see	100	3-L%	NB	Lr	L
125.64	125.87	R:	un yeah	100	F	MB	Br	R
126.15	126.28	L:	de and	*	3-L%	*	*	S
126.28	128.81	L:	sono ano: tyuugoku-ni iku kikkake-n nat-ta-no-ga uh uh China-to go opportunity-DAT become-PAST-N-NOM	*	3-L%	*	*	S
128.81	130.90	L:	tyuutaa-o yat-te-ta-[tte-yuu-koto-[na-n-desu-kedo tutor-ACC do-PAST-QP-thing-COP-N-POL-but	*	3-H%	SB	e	L
129.86	130.19	R:	[un yeah	100	F	MB	B	R
130.43	130.70	R:	[un yeah	100	F	*	B	R
130.86	131.09	R:	[hai yes	100	3-L%	MB	B	R
130.90	131.09	L:	[sore that	*	3-L%	*	*	S
131.09	131.68	L:	daigaku-de university-at	100	3-L%	NB	c	L
132.05	132.62	R:	soo-desu yes-POL	100	3-L%	AB	Lr	L

(*continued*)

Table 4. (*continued*)

Start	End	Sp	Transcript	IPU	IU	CU	PU	UU
132.88	133.31	L:	[(D n)<?> n-	100	D	FB	f	F
132.90	133.14	R:	[un yeah	100	F	MB	Br	R
133.62	134.80	R:	ano: daigaku-de uh university-at	100	3-L%	*	*	S
135.00	136.25	R:	ano daigaku-ttyuu-ka uh university-QP-Q	100	3-L%	WB	*	S
136.53	136.92	R:	(D s) soo y- yes	*	3-L%	NB	*	S
136.94	138.21	R:	ano: tanom-are-te uh ask-PASS-CP	100	3-L%	WB	n	L

4. Characteristics of the proposed utterance-units

In this section, we explore some characteristics of our utterance-units, focusing particularly on unit duration and syntactic property as well as hearers' responses.

4.1 Unit duration and syntactic property

4.1.1 *Purpose*

The aim of this subsection is to examine the prosodic and syntactic properties of our utterance-units and to make clear what kind of units they are. We first analyze the distribution of the durations of SUUs, and show how they are related to units of speaker's speech planning. We then analyze the distribution of the word classes of the last words in SUUs, and discuss its implication with respect to turn-construction.

4.1.2 *Data*

For the dialog data described in Section 2.1, 3,151 SUUs (Chiba corpus: 1,716; CSJ: 1,435) and 1,892 LUUs (Chiba corpus: 1,168; CSJ: 724) were identified by using the procedures shown in Figure 3. Of these data, only those LUUs labeled "L", as well as the SUUs contained in them, were used in the current analysis.

4.1.3 *Results and discussion*

Table 5 shows the 0%, 25%, 50%, 75%, and 100% percentiles of the durations of the SUUs and LUUs in the Chiba corpus and CSJ dialogs. The distributions for the SUUs were relatively narrow, with the inter-quantile ranges (IQRs) being 0.64 s for the Chiba corpus dialogs and 0.68 s for the CSJ dialogs, compared with those for the LUUs, whose IQRs were 1.46 s (Chiba corpus) and 3.19 s (CSJ). In addition, the medians for the SUUs in the two corpora were in accordance with each other, at about 0.7 s. These findings suggest that SUUs may reflect some cognitive process inside the speaker (plausibly speakers' speech planning), that functions uniformly across speech situations. This reminds us of Chafe's notion of *idea units* (Chafe, 1994), realized with a single, coherent intonation contour.

Table 5. Durations (in sec) of short and long utterance-units

	N	0%	25%	50%	75%	100%
Chiba						
SUUs	1154	0.044	0.420	0.714	1.061	4.867
LUUs	617	0.166	0.632	1.107	2.089	15.280
CSJ						
SUUs	1063	0.036	0.445	0.734	1.126	3.101
LUUs	374	0.144	0.852	1.966	4.042	22.430

To look more closely at the duration of SUUs, the data for the SUUs shown in Table 5 were broken down into four sub-sets according to their locations in LUUs, as shown in Table 6. "Initial", "Medial", and "Final" correspond to SUUs located at the initial, medial, and final locations in LUUs, respectively, and "Single" corresponds to any SUU that is coextensive with an LUU. The SUUs at LUU boundaries were longer than the medial SUUs. The durations of the final and single SUUs were longer than those of the initial and medial SUUs. Furthermore, the convergence of the distributions between the two corpora was evident in the medial SUUs, the IQRs being about 0.65 s and the medians being about 0.65 s.

Table 6. Durations of short utterance-units (s) relative to their locations in LUUs

	N	0%	25%	50%	75%	100%
Chiba						
Initial	240	0.050	0.251	0.473	0.804	3.329
Medial	297	0.061	0.364	0.635	1.005	3.647
Final	240	0.044	0.705	0.947	1.389	3.404
Single	377	0.166	0.486	0.730	1.068	4.867
CSJ						
Initial	210	0.054	0.326	0.572	0.952	3.101
Medial	479	0.036	0.400	0.677	1.058	2.607
Final	210	0.094	0.677	1.014	1.457	3.046
Single	164	0.144	0.540	0.811	1.098	2.327

Predictably, the effect of location was also observed in the syntactic property of SUUs. Table 7 shows the top three word classes of the last words in SUUs relative to their locations in LUUs. The final SUUs, and, hence, the LUUs containing them, often ended with final or conjunctive particles or auxiliary verbs (about 80% of the time), which is an expected feature of spoken Japanese utterances (see Section 2.2). For the medial SUUs, on the other hand, case markers were the most frequent word class appearing at SUU boundaries, although their usage rate was not prominent. It is said that turn-construction in Japanese is advanced in an incremental fashion (Tanaka, 1999); case markers progressively project the turn's shape, and utterance-final elements, such as auxiliary verbs and final particles, are placed after the utterance-final predicate and thereby mark a possible completion point of the turn. In this respect, SUUs are building blocks for basic units of interaction, which are realized as LUUs. This perspective is similar to that underlying the idea of *turn-constructional units (TCUs)* (Sacks, Schegloff, & Jefferson, 1974; Schegloff, 1996).

Table 7. Top three word classes* of the last words in SUUs relative to their locations in LUUs

		#1		#2		#3	
Chiba							
Initial	ADV.	(16.7%)	AP	(11.7%)	CN	(11.7%)	
Medial	CM	(15.5%)	ADV.	(10.8%)	CP	(9.8%)	
Final	FP	(42.5%)	CP	(21.7%)	AUX.	(13.3%)	
Single	FP	(32.6%)	AUX.	(16.4%)	ADV.	(10.6%)	
CSJ							
Initial	CONJ.	(17.1%)	ADV.	(14.8%)	CM	(12.9%)	
Medial	CM	(17.5%)	ADV.	(11.7%)	TM	(10.6%)	
Final	FP	(40.0%)	CP	(25.7%)	AUX.	(14.3%)	
Single	FP	(46.3%)	ADV.	(16.5%)	AUX.	(15.9%)	

* CM: case marker, TM: topic marker, AP: adverbial particle, CP: conjunctive particle, FP: final particle, AUX.: auxiliary verb, CN: common noun, ADV.: adverb, CONJ.: conjunction.

4.2 Hearers' responses

4.2.1 *Purpose*

The aim of this subsection is to examine how an utterance-unit being produced by a speaker is treated by other participants, by analyzing the timing of hearers' responses to SUUs and LUUs. We suppose that LUUs constitute basic units of interaction, and, thus, predict that speaker transition would be localized at LUU boundaries. We also suppose that SUUs are not only units of speaker's speech planning but also units of hearer's understanding. Thus, we predict that boundaries of SUUs would provide opportunities for backchanneling and expressive response tokens, which are considered as signals of hearer's understanding and change of hearer's mental state.

4.2.2 *Data and annotation*

In order to distinguish some particular patterns of speaker transitions, the following turn-transition tags were assigned to the LUU data used in Section 4.1 based on a categorization of the ways that conversation progresses, and on turn-taking rules.

First, each dialog was segmented into several chunks, each of which was classified into either a "turn-by-turn" stage or a "telling" stage. In a turn-by-turn stage, utterances are produced in turn by two or more speakers, following the turn-taking system (Sacks et al., 1974), whereas in a telling stage, a single speaker telling a story or giving an explanation exclusively keeps a turn, others supporting his/her multi-unit turn as recipients. Next, for each LUU at a turn-by-turn stage, its antecedent unit was identified by making reference to the time information and the content of the utterance, and the current unit was classified into three types, 1a, 1b, and 1c, according to the turn-taking rules being employed (Sacks et al., 1974):

1a. The current speaker has been selected as next speaker by means of a next-speaker-selection technique, utilized by the speaker of the antecedent unit, such as the affiliation of an address term or a gaze at one party to a class of utterances such as question, request, etc.
1b. The current speaker has selected himself/herself as next speaker, being the first to start a new turn.
1c. The speaker of the antecedent unit has continued his/her turn.

Continuation of a telling sequence by the primary speaker at a telling stage was separately labeled as "s". The annotation was conducted by one of the authors. The question of whether or not the current unit was properly launched at the transition-relevance place (TRP) of the antecedent unit did not figure in the annotation, as the timing of turn-takings with respect to the ends of utterance-units (a type of TRP) was meant to emerge from the distribution of labels. In addition to the four cases above, the case where the current unit was backchanneling or expressive response tokens (RTs) was also included in the data.

4.2.3 *Results and discussion*

Figure 4 shows the histograms of transition times between adjacent LUUs relative to turn-transition types. The ratios of speaker-change types (1a and 1b) to speaker-continuation types (1c and s) in the Chiba corpus and CSJ dialogs were 45.7% (376:447) and 45.9% (214:252), respectively. This means that about a half of the LUUs were accompanied by other participants' start of a new turn.

In the data for types 1a and 1b, we found that the peaks of the distributions were all located at −200 ~ 0 ms or 0 ~ 200 ms and that 95% of the data fell within the range of about 1.5 s in the 1a data (Chiba corpus: 1.4 s; CSJ: 1.6 s) and the range of about 2.3 s in the 1b data (Chiba corpus: 2.4 s; CSJ: 2.2 s). These values contrasted with that in the RT data, which was about 9 s (Chiba corpus: 9.0 s; CSJ: 9.1 s).

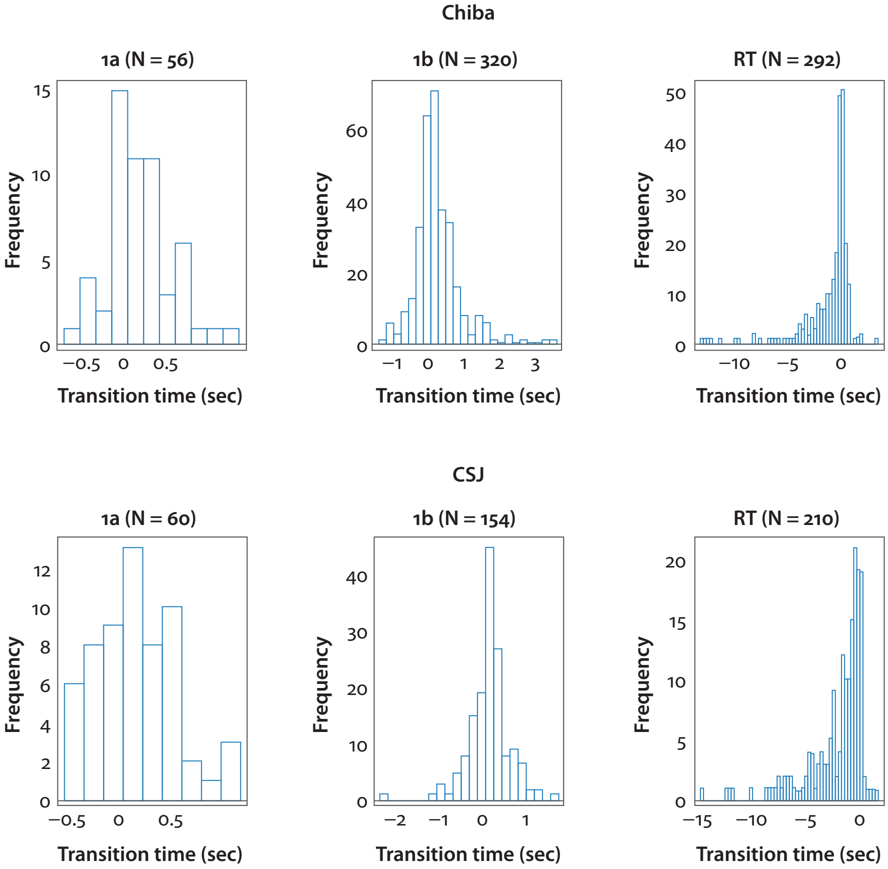


Figure 4. Distributions of transition times between adjacent long utterance-units relative to turn-transition types. The histograms for the speaker-continuation types (1c and s) were omitted. In each histogram, the bin-width is fixed to 200 ms

In order to see the relation of hearers' responses to SUUs, the target SUU in the antecedent unit was also defined for each LUU pair in the data as the last SUU whose ending time was not later than the starting time of the current unit. The solid lines in Figure 5 show the observed distributions of the positions of target SUUs measured from the end of the antecedent unit for the 1a, 1b, and RT data. The broken lines, on the other hand, show the distributions predicted by a model in which the target SUU was selected from each antecedent unit with equal probability. As clearly seen in the 1a and 1b data, virtually all responses occurred at the final SUU in the antecedent unit, although, in theory, earlier SUUs also could have been the target. A dramatic difference, however, was observed in the RT data, where the observed and the predicted distributions were rather similar.

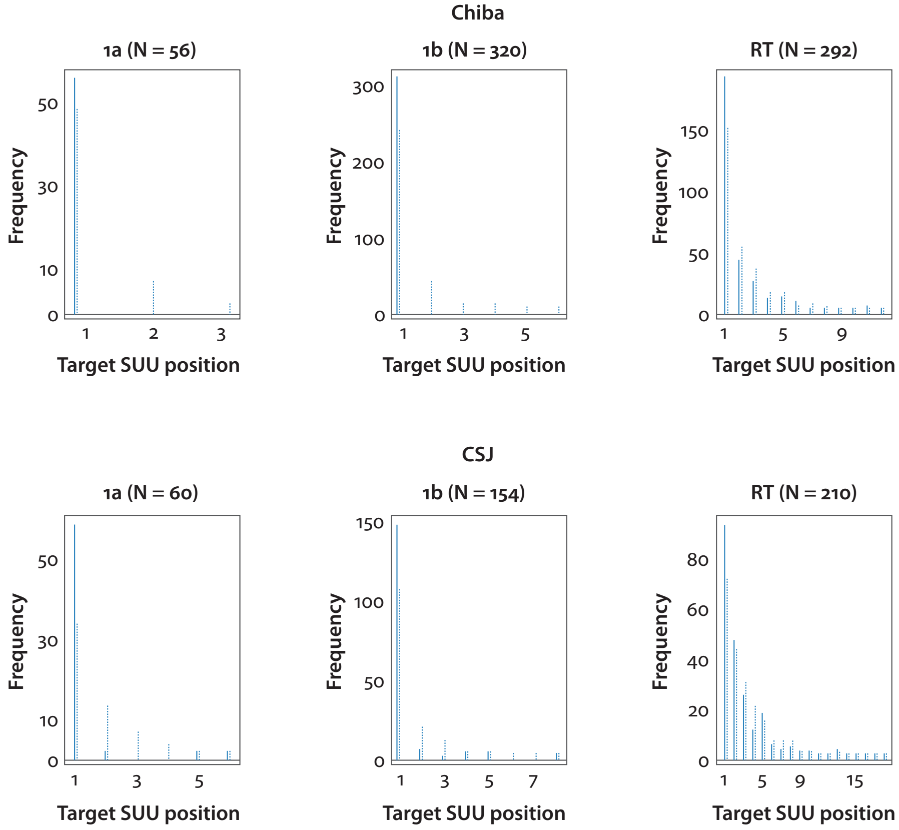


Figure 5. Distributions of target SUU positions measured from the end of the antecedent unit relative to turn-transition types

In Figure 5, solid lines represent the observed distributions, whereas broken lines represent the distributions predicted by a random model in which the target SUU was selected from each antecedent unit with equal probability. In sum, the timing of turn-taking was localized at LUU boundaries, suggesting the adequacy of LUUs as units of interaction. In contrast, the chance of eliciting a response token was nearly equal for all SUUs contained in an LUU, suggesting that SUUs may be well suited for units of hearer's understanding and change of hearer's mental state.

5. Extensions to the scheme

In Section 2, four types of units, IPUs, IUs, CUs, and PUs were identified separately. The statistical analysis of the interrelationship among the four showed that the distributions of these disjunctures could be classified into five groups. Based on this result, in Section 3, we proposed a two-level annotation scheme of utterance-unit in Japanese dialog, SUU and LUU, and examined their characteristics in Section 4.

At this point, let us consider two questions. Is our annotation scheme exhaustive enough to extract basic units in dialog? And, is it adequate to assume a hierarchical relationship between SUUs and LUUs? The first point addresses whether any other kind of boundaries should be identified in dialog data, and the second asks whether it is appropriate to presuppose a subsumptive relationship between SUUs and LUUs. In this section, we examine these two questions, and, at the end, extend our two-level annotation scheme.

5.1 Interactional disjuncture

The first question is whether our annotation scheme of SUU and LUU is adequate to capture various phenomena in dialog. As we have shown in Section 4, one of the important characteristics of dialog is that it involves massive interactions between speakers and hearers. There are some cases that require types of disjuncture other than acoustic/prosodic and syntactic/pragmatic ones if our purpose is to use annotation for an observationally adequate analysis of interaction.

Table 8. Example of an increment after a possible completion point.

	Start	End	Sp	Transcript
→	112.42	113.18	A:	sasuga-da-yo great-COP-FP
	113.41	113.79	C:	un yeah
	114.03	114.20	A:	ne isn’t he?

In Table 8, participant A once completes her utterance *sasuga-da-yo* ‘he is great’ with a sentence-final particle *yo*, and waits for the hearer’s response. After C gives a response with *un* ‘yeah’, A immediately adds another sentence-final particle *ne* to re-complete her utterance. The point after *sasuga-da-yo* can be regarded as a *possible completion point* (Sacks et al., 1974), where another participant can initiate a new turn (and C actually does). The original scheme, however, does not identify a boundary after *sasuga-da-yo*, because incorporating the subsequent word *ne* will yield a complete syntactic unit.

To extend the boundary classification rule, we introduce a new criteria to identify interactional disjuncture. If the speaker stops his/her utterance and waits and sees the hearer's response, that boundary is labeled "L" as a point where an interaction can occur. These criteria can be applied to the cases where the speaker stops his/her utterance even in the middle of a syntactic structure. In Table 9, the participants are talking about the payment for their training camp, and A and B ask C whether he can pay 10,000 yen. When C starts his turn with *itioo oya-ni* 'as one recourse, from my parents', which is syntactically incomplete at that moment, A and B simultaneously respond with *a:* 'ah', while C continues his utterance to bring it to a syntactic completion. The point after *itioo oya-ni* can also be regarded as a possible completion point, and, thus, should be identified as an interactional disjuncture.

Table 9. Example of a possible completion point before syntactic disjuncture.

	Start	End	Sp	Transcript
	283.63	284.22	A:	kane haraen-no money can.pay-Q
	284.31	285.04	A:	ano: gassyuku-dai uh training.camp-cost
	((7 lines omitted))			
	289.73	289.83	C:	a ah
→	289.83	290.74	C:	itioo oya-ni as one recourse parents-DAT
	291.30	291.97	A:	a [: ah
	291.48	292.50	B:	[a: ah
	291.53	292.87	C:	[maikai mora-tte-masu-kedo every.time get-CP-POL-but

Another example of interactional disjuncture is shown in Table 10. B starts her turn with *nanka:* 'it's something like' in the middle of A's turn, and right after this word, B responds to A with *soo* 'right'. B's original utterance, which has been started with *nanka:,* is now interrupted by herself, although she resumes it, after the response to A, by saying *namae-ga tigau* 'the name is different'. Since *soo* constitutes a response, which should be identified as an independent utterance unit, the part of the utterance up to this point should also be identified as a separate utterance unit: a fragment in this case.

The boundary classification rule for these cases would be as follows: If the speaker responds to other participant in the middle of an utterance being constructed, the point right before the response is identified as an interactional disjuncture and labeled "L".

Table 10. Example of a response in the middle of an utterance.

	Start	End	Sp	Transcript
	76.67	79.07	A:	kafe conpaana-toka ano [hen-ga so[o naru-ka-na= cafe con.panna-or.anything that area-NOM concerned be-Q-FP
	78.06	78.35	B:	[un yeah
→	78.35	78.99	B:	[nanka: something.like
	78.99	79.37	B:	=soo right
	79.77	80.62	B:	namae-ga tigau name-NOM different

5.2 Mismatch between short and long utterance-unit boundaries

In prescribing the utterance-unit identification rules shown in Figure 3, we have assumed hierarchical relationship between SUUs and LUUs; in other words, we have presupposed that boundaries labeled "L" not only constitute LUU boundaries but also subsume SUU boundaries. The second question in this section is whether this assumption is valid or not. Is there any case where a mismatch between SUU and LUU occurs?

There were a few cases where an LUU boundary did not exhibit the property of an SUU boundary, that is, acoustic or prosodic disjuncture. Among the 991 LUUs used in Section 4.1, 64 instances (Chiba corpus: 52 (= 8%); CSJ: 12 (= 3%)) did not share any properties of SUU boundaries. The majority of these instances could be classified into the following patterns. In the following examples, # indicates a mismatch boundary, that is, an LUU boundary that lacks the property of an SUU boundary.

1. Postposed constructions (18 cases):
   The mismatch boundary is immediately followed by a postposed element: for example, *ii-too-no sit-teru:#ano koohii-meekaa* 'Do you know the one in building E?#that coffee machine'.
2. Turn prefaces (6 cases):
   The LUU in question is a preface to the body of the speaker's turn, projecting the continuation of his/her turn across the mismatch boundary: for example, *itsumo omou-n-da-kedo#X-san-tte Y-san-ni-taisi-te tyotto-sa: kekkoo: yuu-yo-ne* 'I always think about this, but#Ms. X just says a lot to Mr. Y, doesn't she?'.

3. Lexical response tokens (10 cases):
   The LUU in question is a lexical response token *soo* or *soo-desu(-ne)*, which is immediately followed by a substantial utterance by the same speaker, yielding a resumptive opener (Clancy et al., 1996): for example, (responding to a question "Is it like a cup?") *soo#de sore-no repurika-to-ka ut-teru-tte* 'Right#and they sell its replica or something'.
4. Repeats of predicates (4 cases):
   The LUU in question is repeated immediately afterward for the purpose of emphasis, etc.: for example, *it-teru#it-teru* 'It's hissing#hissing'.

At these mismatch boundaries, other participants rarely started their new turn; the rate was 27% in the Chiba corpus dialogs and 0% in the CSJ dialogs. This fact might suggest that the speaker sometimes utilizes some kind of technique to "rush through" the completion point in order to keep his/her turn (Schegloff, 1987). Although there remains much to be discussed concerning this issue (see Den et al., 2010, for more discussion), it is obvious that mismatch between SUU and LUU provides a rich source for the study of interaction.

5.3 Revised scheme

As shown in the previous two subsections, our current scheme has room for revision mostly from the view point of interactional phenomena. Since it is worth allowing mismatches between SUU and LUU, we do not presuppose hierarchical relationship between SUUs and LUUs, and identify the two units separately. Figure 6 shows a revised scheme. Different label sets are used for SUUs and LUUs; the former represented by lower case letters, "f" and "s" and the latter by upper case letters, "F", "R", and "L". Note that separate application of the rules for the two units enables us to obtain LUUs whose boundaries are not SUU boundaries.

The extended scheme has been applied to the public version of the *Chiba Three-Party Conversation Corpus*,[1] and its application is planned for the *Corpus of Everyday Japanese Conversation* (Koiso et al., 2018), which is being developed at the National Institute for Japanese Language and Linguistics.

[1] <http://research.nii.ac.jp/src/en/Chiba3Party.html>

[Unit identification/classification rule for short utterance-unit (SUU)]

Apply the following rules in this order at every word boundary:

1. If the tokens so far constitute a fragment/filler or a backchanneling/expressive interjection, mark the boundary with “f”;
2. Else if the current boundary is an acoustic and/or prosodic disjuncture, which may be expressed by a pause or an intonation break, mark the boundary with “s”;
3. Otherwise, apply these rules at the next word boundary.

Identify the two types of boundaries above as boundaries of short utterance-units.

[Unit identification/classification rule for long utterance-unit (LUU)]

Apply the following rules in this order at every word boundary:

1. If the tokens so far constitute a fragment or an interruption followed by a pause, mark the boundary with “F”;
2. Else if the tokens so far constitute a backchanneling or expressive interjection, including those which function as a reply or an acknowledgment, mark the boundary with “R”;
3. Else if the current boundary is a syntactic, pragmatic, and/or interactional disjuncture, which may be expressed by a clause-unit boundary, a linguistic modality, or a possible interaction between the speaker and the hearer, mark the boundary with “L”;
4. Otherwise, apply these rules at the next word boundary.

Identify the three types of boundaries above as boundaries of long utterance-units.

Figure 6. Rules identifying short and long utterance-units

6. Concluding remarks

In this paper, we have proposed a new scheme to annotate basic speech units in dialog, short and long utterance-units. Our new scheme covers boundaries of units at different levels including acoustic, prosodic, syntactic, pragmatic, and interactional disjunctures. This multi-layered annotation scheme makes it possible to extract basic units to handle dialog data from various aspects of linguistics and related research areas including phonology, syntax, semantics, pragmatics, discourse analysis, conversation analysis, and psycholinguistics.

What nature do SUUs and LUUs have as basic units in dialog? SUUs are identified by acoustic and prosodic boundary such as a pause and an intonation break. This idea of SUUs is consistent with Chafe’s notion of *idea unit*s (Chafe, 1980), or *intonation units* (Chafe, 1994), a notion which is considered important and influential in spoken discourse studies from a cognitive viewpoint. Facts presented in Section 4 suggest a similarity between the ideas of Chafe’s intonation units and our SUUs to the extent that both reflect speakers’ planning process. Chafe (1987) argues that an intonation unit expresses “a single focus of consciousness” of a speaker, which is associated with the speaker’s cognition. We propose that the same property is seen in SUUs: The SUU is a cognitively formulated unit with a single focus of consciousness into a short spurt to be uttered at once in Japanese dialog.

We have also found support for the idea that SUUs are units of hearer’s understanding. Examining hearers’ responses in the dialog data, we showed that SUUs provide opportunities for the hearer to respond to the speaker with brief response tokens. Hearers sometimes display their understanding with various forms of response tokens immediately after SUU boundaries. This means that hearers utilize SUUs effectively as units of their incremental understanding. In this respect, we can say that SUUs are useful in the study of cognitive aspect of spoken dialogs from not only speaker’s but also hearer’s perspective.

On the other hand, syntactic, pragmatic, and interactional boundaries play an important role in recognizing LUUs. As we have shown in Section 5.2, SUUs and LUUs may not always stand in a hierarchical relationship. Still, the majority of LUU boundaries are also acoustic and/or prosodic disjuncture, and LUUs can roughly be seen as units defined by the convergence of prosodic, syntactic, and pragmatic completions. These characteristics of LUUs are parallel to *complex turn-constructional units*, which have been proposed as basic unit of interaction in the conversation analysis literature (Ford & Thompson, 1996). Some characteristics of LUUs presented in Section 4 also suggested the adequacy of LUU as a unit of interaction. In these respects, we can say that LUUs are useful in the study of communicative and interactional aspects of spoken dialogs.

To promote these lines of research, we are developing an annotation scheme specifying the *functions* of utterance-units. For the study of cognitive aspect, the information status of discourse elements as “given” and “new”, etc., may be of use. We are investigating approaches for annotating such functional information on SUUs (Nakagawa & Den, 2012). For the study of aspects of communicative/interactional behavior, on the other hand, it is fundamental to represent the structures of turns and the actions performed therein. We are developing an annotation scheme for LUUs to represent dialog acts, which is applicable not only to dialogs conducted in experimental or artificial settings but also to everyday, real-life conversations.

Acknowledgements

This study has been supported by the Grant-in-Aid for Scientific Research (# 20320057, 23320081, 24520523, and 16H03426), and NINJAL’s Collaborative Research Project “A Multifaceted Study of Spoken Language Using a Large-scale Corpus of Everyday Japanese Conversation”.

This paper is an extended version of the paper presented at the 7th International Conference on Language Resources and Evaluation (LREC 2010) (Den et al., 2010). The authors would like to thank the other co-authors of the original paper.

The following glosses are used in this paper

NOM	nominative case marker	PAST	past tense
ACC	accusative case marker	POL	politeness marker
DAT	dative case marker	CP	conjunctive particle
TOP	topic marker	FP	final particle
COP	copula	QP	quotative particle
N	nominalizer	Q	question marker
PASS	passive voice		

References

AMI Project (2005). *Guidelines for dialogue act and addressee annotation* (Version 1.0).

Beckman, M., & Ayers, G. (1994). *Guidelines for ToBI labeling*. Columbus, OH: Ohio State University.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English*. London: Longman.

Chafe, W. L. (1980). The deployment of consciousness in the construction of narrative. In W. L. Chafe (Ed.), *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production* (pp. 353–387). Norwood, NJ: Ablex.

Chafe, W. L. (1987). Cognitive constraints on information flow. In R. S. Tomlin (Ed.), *Coherence and grounding in discourse* (pp. 21–51). Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.11.03cha

Chafe, W. L. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago, IL: University of Chicago Press.

Clancy, P. M., Thompson, S. A., Suzuki, R., & Tao, H. (1996). The conversational use of reactive tokens in English, Japanese, and Mandarin. *Journal of Pragmatics*, 26, 355–387. https://doi.org/10.1016/0378-2166(95)00036-4

Den, Y., & Enomoto, M. (2007). A scientific approach to conversational informatics: Description, analysis, and modeling of human conversation. In N. T. (Ed.), *Conversational informatics: An engineering approach* (pp. 307–330). Hoboken, NJ: John Wiley & Sons. https://doi.org/10.1002/9780470512470.ch17

Den, Y., Koiso, H., Maruyama, T., Maekawa, K., Takanashi, K., Enomoto, M., & Yoshida, N. (2010). Two-level annotation of utterance-units in Japanese dialogs: An empirically emerged scheme. In *Proceedings of the 7th international conference on language resources and evaluation (LREC 2010)* (pp. 2103–2110). Valletta, Malta.

Den, Y., Yoshida, N., Takanashi, K., & Koiso, H. (2011). Annotation of Japanese response tokens and preliminary analysis on their distribution in three-party conversations. In *Proceedings of the 14th Oriental COCOSDA (O-COCOSDA 2011)* (pp. 168–173). Hsinchu, Taiwan.

Du Bois, J. W., Shuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of discourse transcription. In J. A. Edwards & M. D. Lampert (Eds.), *Talking data: Transcription and coding in discourse research* (pp. 45–89). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 134–184). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620874.003

Gardner, R. (2001). *When listeners talk*. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.92

Iwasaki, S. (1993). The structure of the intonation unit in Japanese. In S. Choi (Ed.), *Japanese/Korean linguistics* (Vol. 3, pp. 39–53). Stanford, CA: CSLI.

Iwasaki, S., & Ono, T. (2002). "Sentence" in spontaneous spoken Japanese discourse. In J. Bybee, & M. Noonan (Eds.), *Complex sentences in grammar and discourse: Essays in honor of Sandra A. Thompson* (pp. 175–202). Amsterdam: John Benjamins. https://doi.org/10.1075/z.110.10iwa

Koiso, H., Den, Y., Iseki, Y., Kashino, W., Kawabata, Y., Nishikawa, K., Tanaka, Y., & Usuda, Y. (2018). Construction of the corpus of everyday Japanese conversation: An interim report. In *Proceedings of the 11th international conference on language resources and evaluation (LREC 2018)* (pp. 4259–4264). Miyazaki, Japan.

Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A., & Den, Y. (1998). An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese Map Task dialogs. *Language and Speech*, 41, 295–321. https://doi.org/10.1177/002383099804100404

Maekawa, K. (2003). Corpus of spontaneous Japanese: Its design and evaluation. In *Proceedings of ISCA and IEEE workshop on spontaneous speech processing and recognition (SSPR 2003)* (pp. 7–12). Tokyo.

Maekawa, K., Kikuchi, H., Igarashi, Y., & Venditti, J. J. (2002). X-JToBI: An extended J ToBI for spontaneous speech. In *Proceedings of the 7th international conference on spoken language processing (INTERSPEECH 2002)* (pp. 1545–1548). Denver, CO.

Maruyama, T., Horn, S. W., Russell, K. L., & Frellesvig, B. (2017). On the multiple clause linkage structure of Japanese: A corpus-based study. In N. Kazashi & M. Mariotti (Eds.), *New steps in Japanese studies: Kobe university joint research* (pp. 131–154). Venice: Ca' Foscari.

Maruyama, T., Takanashi, K., & Yoshida, N. (2010). An annotation scheme for syntactic unit in Japanese dialog. In *Proceedings of DiSS-LPSS joint workshop 2010: The 5th workshop on disfluency in spontaneous speech, and the 2nd international symposium on linguistic patterns in spontaneous speech* (pp. 51–54). Tokyo.

Meteer, M., Taylor, A., MacIntyre, R., & Iyer, R. (1995). *Dysfluency annotation stylebook for the switchboard corpus*. Philadelphia, PA: University of Pennsylvania.

Nakagawa, N., & Den, Y. (2012). Annotation of anaphoric relations and topic continuity in Japanese conversation. In *Proceedings of the 8th international conference on language resources and evaluation (LREC 2012)* (pp. 179–186). Istanbul, Turkey.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 696–735. https://doi.org/10.1353/lan.1974.0010

Schegloff, E. A. (1987). Recycled turn beginnings: A precise repair mechanism in conversation's turn-taking organization. In G. Button & J. R. E. Lee (Eds.), *Talk and social organisation* (pp. 70–85). Clevedon: Multilingual Matters.

Schegloff, E. A. (1996). Turn organization: One intersection of grammar and interaction. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 52–133). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620874.002

Takanashi, K., Maruyama, T., Uchimoto, K., & Isahara, H. (2003). Identification of "sentences" in spontaneous Japanese – Detection and modification of clause boundaries –. In *Proceedings of ISCA and IEEE workshop on spontaneous speech processing and recognition (SSPR 2003)* (pp. 183–186). Tokyo.

Tanaka, H. (1999). *Turn-taking in Japanese conversation: A study in grammar and interaction*. Amsterdam: John Benjamins.

Venditti, J. J. (1994). *Japanese ToBI labelling guidelines*, 50. Columbus, OH: Ohio State University.

CHAPTER 6

The pragmatic analysis of speech and its illocutionary classification according to the Language into Act Theory

Emanuela Cresti
University of Florence – LABLITA

According to the Language into Act Theory, reference units in speech have a pragmatic nature: they correspond to the activation of sensory-motor schemas leading to the performance of different speech acts. Our background is the affective and psychic motivations of the *Human Birth Theory* (Fagioli, 1971), compatible with recent theories of *Embodied Cognition*. Identification and classification of speech acts rest on corpus-based research. Speech activity is encoded by prosody, that conveys *utterance boundaries, illocutionary force* and *information structure*. The utterance nucleus is the *Comment*, responsible for illocutionary force. We illustrate the methodology for the induction of illocutions from corpora and detail pragmatic and prosodic features which allow classifying illocutionary types. A case study is presented for four original illocutions (self-conclusion, assertion taken for granted, ascertainment, evidentiality assertion).

Keywords: speech activity, illocutions, spoken corpora, prosody, information structure

1. Premises

1.1 Introduction to Language into Act Theory

Within the tradition of Austin (1962), Language into Act Theory (L-AcT; Cresti, 2000) assumes that three acts accomplished simultaneously in a speech act (locution, illocution, perlocution) correspond to an organic system governed by specific rules. In brief it may be described with the following schema: The speech activity finds its origins in a mental/affective representation, which is a reaction to an external input (perlocutionary act) and is transformed into a linguistic action schema toward the addressee, conventionally codified in every culture as a pragmatic issue (illocutionary act). Through the interface of prosody, the latter functionally

https://doi.org/10.1075/scl.94.06cre
Additional files available at http://doi.org/10.1075/scl.94.06cre/audio

determines the linguistic chunks of speech, that is, the semantic/syntactic islands that are characterized according to the specific language (locutionary act).

With respect to its psychological foundation the L-AcT approach can be traced to philosophical assumptions which have been conceived in Italy by the psychiatrist Fagioli (2010, 2011, 2012) in his *Human Birth Theory*.[1] At birth, a primitive neuro-psychophysiological condition (*pulsion*) produces the activation of the cerebral cortex through the stimulation of light energy. The fusion between the pulsion, which for defence would lead to the annulment of the hostile non-human world (light, cold, noise), and the biological vitality triggers a reaction by developing the capability to imagine. The first thought, indeed, is a fantasy activity of an undefined mental image. The previous intrauterine condition is turned into an internal image, a memory-fantasy of the sensation had before, giving rise to our way of thinking. The foundation of our thought is substantiated by images and an irrational character, existing as an ideational/affective reaction to the external inputs (mostly human dynamics) and lasting for a lifetime. After about one year the child begins to speak, manifesting thought through language according to affective behavior directed at the addressee.

Within the L-AcT perspective, the ideation found in the speech origin is shaped by the speaker's affect toward the addressee and is physically transformed into a speech act containing a conventional pragmatic value via prosodic devices. Within this model, the linguistic content and its syntactic structure are dependent on pragmatic/affective functions. There is a pragmatic activity at the origin of speech which depends on the speaker's affect toward the addressee.

With regards to the involvement of a pragmatic aspect in speech activation, this seems to have found confirmation in recent lines of research regarding the *embodiment* of cognition (Arbib, 2012). The concept of embodiment is supported by neurobiological research that grounds human cognition in a sensory-motor system, which is shared with motion and perception. For speech, a complex motor system activation is foreseen that depends on that enacted for non-linguistic actions (Egorova, Shtyrov, & Pulvermüller, 2015; Mollo, Pulvermüller, & Hauk, 2016). For instance, the processing of action related utterances involves the activation of the motor system in the brain, leading to the embodied representations of the linguistic meaning. More generally, neuroscientists investigated the systems which underlie both language and sensory-motor faculties, and developed models in which linguistic, cognitive and motor abilities strictly interact. L-AcT's aim is to

[1] For a detailed presentation of Human Birth Theory (HBT) see https://massimofagioli.com/en/human-birth-theory/ and the Wikipedia entry. For recent references see Calesini (2017), Gatti et al. (2012), Giorgini et al. (forthcoming), Maccari, Polese, Reynaert, & Fagioli (2016).

explain what the activity of the speaker is when he is talking, considering it as an organic system in its whole.

The assumption of the pragmatic level as a mandatory passage seems to constitute a marked point of divergence from other research into spoken language. For instance, one of the most known models, that of Chafe (1970), does not consider the necessity of this step in passing from thought to speech and foresees a direct correspondence between ideas and prosodic/linguistic units. The L-AcT conception of speech was developed independently but in parallel with another important avenue of research: the macro-syntactic approach carried out by *Groupe Aixois de Recherche en Syntaxe* (GARS), which was founded by Blanche-Benveniste at the University of Aix-en-Provence (Blanche-Benveniste, 1997, 2003; Blanche-Benveniste, Rouget, Bilger, & van den Eynde, 1990). The GARS's model of syntactic analysis is composed of two levels: the micro-syntactic and macro-syntactic layers, introducing a different kind of perspective for the identification of the reference unit in speech. The macro-syntactic approach has been further developed in recent years by the project Rhapsodie (Lacheret-Dujour, Kahane, & Pietrandrea, 2018), which enriches the macro-syntactic perspective with a set of modular components that include pragmatics, information structure, and prosody. It has led to the identification of the *illocutionary unit* as the maximal entity, with a possible mapping to a maximal prosodic entity (the *intonational period*), approaching L-AcT's research stance. It foresees three different mechanisms (syntactic, macro-syntactic/illocutionary, prosodic) governing the activation of speech, but these are considered independent from one another (Debaisieux, 2013).[2]

Within L-Act the information structure finds its center in the pragmatic accomplishment of a necessary *information unit* known as the *Comment*. The *Comment* may be accompanied by optional components (within the same information structure), forming an *information pattern*. The additional units develop different functions (*Topic, Parenthesis, Appendix, Locutive Introducer, Discourse Markers*).

This conception of the information structure moves away from one of the more popular traditions of nowadays, that of Krifka (Krifka, 2007; Krifka & Musan, 2012). Krifka's approach is grounded in natural logics and finds in the context, that is, the *Common Ground* (Stalnaker, 1999), the conditioning origin of information structure and, finally, speech. In contrast, L-AcT focuses on the subjective initiative of the speaker toward the addressee, reacting to context but not depending on it.

[2] The macro-syntactic research has been accompanied – primarily in the French-speaking world – by neighbouring works that share a similar scientific approach such as the "school" of Neuchâtel and Freiburg (Berrendonner, Béguelin, Avanzi) and that of Louvain (Degand, Simon, Mertens).

The necessary new component in each utterance is always the speaker's affective/pragmatic reaction, expressed by the Comment, and is unpredictable.

With regard to prosody, we agree on many points with Martin (2015), including that prosodic units are not a sort of superficial execution of deep syntactic structures, which the generative tradition maintains. It must be considered, indeed, that prosody has its own rules and conditions, such as delimiting the maximal length of a prosodic unit (7 syllables), preventing clashes deriving from contiguous stressed syllables, the general decreasing f0 trend (i.e., the *declination line*), and a principle of *dependence on the future* between movement characterizing the prosodic units inside a prosodic pattern. But in our opinion, prosody is integrated within the system performing the information pattern, including its illocutionary center, transforming a pragmatic schema into a linguistic object. Within L-AcT, prosody is considered the interface between the illocutionary and locutionary act and constitutes the necessary means of *transducing* the pragmatic conception into a concrete and audible entity. In some sense it can be defined as a projection of the pragmatic activation, but not just this, as prosody finds its origins in the affective intention toward the addressee, motivating speech.[3]

Our research had to face the fundamental problem of identifying *speech reference units*, which are intended to be those entities to which we refer specifically for the linguistic analysis of speech. The question traces back to an extended debate in which specialists working on spoken language concluded that the sentence, defined as a complete and well-formed syntactic configuration, was not an adequate entity on which to ground the analysis of speech. The search for a speech reference unit is one of looking for entities that can be traced to syntactic, semantic, pragmatic or prosodic characteristics or to their correlations in order to be able to generate linguistically meaningful units. For this reason, the *turn* is not considered, though it is in fact a natural unit for the identification of speech in terms of silence and change of speaker voice; a turn may last a few milliseconds to a few minutes and cannot be formalized into a linguistic entity.[4]

[3] Prosody develops many functions, since it is the first attentional input of newborns and constitutes their first device to signal and communicate to other human beings. It accompanies our entire linguistic acquisition and grounds our competence formation. However, given that it is rooted in our emotional and affective assets it also represents the device for expressing attitudes, moods, and personal connotations.

[4] A different line of research known as Interactionism must also be mentioned (Barth-Weingarten, Reber, & Selting, 2010; Couper-Kuhlen, 2004). The *dialogic turn* has been elected as the natural unit of speech. The turn should solve the question of the actual identification of the reference unit, given the easy way in which it may be identified in speech flow via the speakers' change of voice. However, given the difference in the linguistic contents of the turns, this choice produces

Thus, many proposals have been put forth, such as: the *clause*, intended as a phrasal unit with different possible syntactic fulfilments (Miller & Weinart, 1998), the *C-unit* which may correspond to a noun phrase (Biber, Johansson, Leech, Conrad, & Finegan, 1999), the Basic Discourse Unit (*BDU*), identified through a correlation between syntax and prosody (Degand & Simon, 2009), and the *utterance*, identified through the correlation between pragmatics and prosody (Cresti, 2000, 2018).

In fact, corpus-driven research led to the discovery of two types of pragmatic constructs, both identified by prosody: the *utterance,* being the counterpart of the accomplishment of a single speech act, and the *stanza*, forming a sequence of weak Comments, outside any program as good as it gets, continuing until the end of the thought flow.[5] A stanza does not correspond to the sum of utterances but is a new type of reference unit composed of two or more Bound Comments, each of which may in turn be supported by other information units (forming sub-patterns). The stanza approaches the definition of a reference unit proposed by Chafe – the Idea Unit (Chafe, 1980) –, since its pragmatic value is rather low, following the expression of a stretch of thought more than an interactive exchange. It is not by chance that stanzas occur especially in monologic and formal texts.[6] Prosody delimits the boundaries of reference entities, both utterances and stanzas, and it is the necessary interface between the illocutionary and locutionary act. In conclusion, our approach focuses in on the pragmatic aspect of the utterances and stanzas and on their information patterning, which is performed by prosody.

1.2 The corpus-based analysis of spontaneous speech

In accordance with the Austinian tradition of studies, L-AcT considers the utterance to be the *minimal linguistic entity* that is *pragmatically interpretable* and the primary reference unit for speech analysis. The aim of L-AcT is to ground

other relevant questions. Interactional researchers propose kinds of sub-turn or virtual turn to overcome this difficulty, but these entities end up practically coinciding with the syntactic clause.

[5] The distinctions of weak and strong assertive illocutionary types are based on the different degrees of relevance of the semantic content in the utterance, the (speaker's) commitment to the content's truth, and the degree of the speaker's involvement with respect to the addressee. For details see Section 5.2.

[6] The quantitative values for the two reference units is quite different, according to Italian IPIC (Panunzi & Gregori, 2011) the utterance records at approximately 90% and the stanza 10% for the total number of terminated prosodic sequences.

the systematic analysis of spoken corpora in speech act theory.[7] To this end a corpus-based methodology has been developed, and its main innovation with respect to Austin's work is in considering the spoken activity (the illocutionary force and the information structure) as manifested through prosodic devices. On this assumption, prosodic processing is a necessary step for the pragmatic analysis of speech and is based on the prosodic identification of the utterance and its information pattern in the flow of speech.

The identification of reference units within speech flow via prosody is achieved through the detection of those prosodic breaks which are perceived as being terminal. This approach is the result of a long analysis carried out by the LABLITA[8] team over the last 30 years, in which they verified a systematic correspondence between a stretch of speech ending with a *terminal prosodic break* and the accomplishment of an illocutionary force (Cresti, 2000; Cresti & Moneglia, 2005; Moneglia, 2006; Moneglia & Cresti, 2006).

Classic studies on prosody have always highlighted the fact that sentences end with a terminal profile (Crystal, 1975; Karcevsky, 1931). This property is used in L-AcT as a heuristic for perceptually determining utterance boundaries in the speech flow and traces back to the IPO (Institute for Perception Research), which stresses the perceptual relevance of intentionally performed prosodic cues ('t Hart, Collier, & Cohen, 1990). Indeed, competent speakers perceive breaks easily and can also distinguish between prosodic boundary types with a terminal value (terminal breaks) and boundaries which indicate that the utterance will continue (non-terminal breaks; Swerts, 1997).[9]

These cues are so prominent that little training is required for the accurate identification of breaks. Furthermore, the practice of perceptual recognition by

[7] L-AcT has been in development since the early eighties and has been tested extensively through the collection and annotation of spoken Romance corpora: the LABLITA corpus (Cresti, Moneglia, & Panunzi, 2018); C-ORAL-ROM (Cresti & Moneglia, 2005); C-ORAL-BRASIL (Raso & Mello, 2012); Cor-DiAL (Nicolás Martínez, 2012). L-AcT has been used as a basis for the cross-linguistic comparison of Information Structure in spontaneous speech (IPIC Data Base: Mittmann & Raso, 2012; Moneglia & Cresti, 2015; Nicolas & Lombán, 2018; Panunzi & Gregori, 2011; Panunzi & Mittmann, 2014). The framework has also been applied by LEEL team to a comparable American English corpus (Cavalcante & Ramos, 2016), taken from the Santa Barbara Corpus (Du Bois, Chafe, Meyer, & Thompson, 2000).

[8] LABLITA is the acronym for the Laboratory of Italian Linguistics founded by Cresti and Moneglia in 1985 within the Department of Literature and Philosophy at the University of Florence. It archives important written and spoken Italian corpora and carries out research into semantics, pragmatics, information structure and prosody.

[9] According to the LABLITA transcription format they are marked respectively with a double slash (//) and a single one (/).

native speakers has been adopted successfully in the transcription and annotation of large corpora (Buhmann et al., 2002; Cheng, Greaves, & Warren, 2008; Cresti & Moneglia, 2005; Du Bois, Cumming, Schuetze-Coburn, & Paolino, 1992; Izre'el, 2005; Izre'el & Mettouchi, 2015; Raso & Mello, 2012). This perceptual criterion grounding the annotation of prosodic breaks has been validated within the Dutch Corpus (Buhmann et al., 2002) and within C-ORAL-ROM and C-ORAL-BRASIL (Danieli et al., 2004; Mello, Raso, Mittmann, Vale, & Côrtes, 2012; Moneglia, Raso, Mittmann, & Mello, 2010; Raso & Mittmann, 2009).

The parsing of speech via prosody does not end with the demarcation of terminated sequences; an utterance's text may be further segmented into prosodic units. The resultant substructure is interpreted as correlating with the information structure of the utterance, which has been called information packaging by Chafe (1970). With this viewpoint, L-AcT assumes that every utterance corresponds to an *Information Pattern* which may be composed of many information units, systematically demarcated by *non-terminal prosodic breaks* (Izre'el, 2005; Moneglia & Cresti, 2006; Raso, 2014; Swerts, 1997; Swerts & Geluykens, 1993).

An Information Pattern may be *simple*, which is to say composed of only one information unit. In C-ORAL-ROM this makes up a large percentage of the utterances: more than 35%. In these cases, the illocutionary information is conveyed by a single information unit, that is, *Comment*. From this we can observe that the Comment is necessary and sufficient for fulfilling an utterance. Moreover, one and only one prosodic unit type, called the *root* (following the IPO system), may perform the Comment and conveys the illocutionary value.

However, an Information Pattern may also be *compounded*, in which case the Comment is accompanied by other, optional units, though it is still the first that conveys the illocutionary value of the utterance. Each of the other information units is performed by a prosodic unit type and conveys a specific information function with respect to the Comment (Moneglia & Raso, 2014).

1.3 The empirical research

The organization of the LABLITA corpus entails a set of variation parameters which are considered relevant for representing spontaneous speech (Biber, 1988; Mello, 2014) and, specifically, its dia-phasic variation, so ensuring the pragmatic representation of speech and specifically of its illocutionary characterization.

Research into illocution and its classification has always been a challenge (Kempson, 1977; Leech, 2014; Sbisà, 1989; Sbisà & Turner, 2013). Beyond the well-known illocutionary types such as assertion, order, question – reducing the illocutionary variety to the syntactic typologies of sentence: *declarative, jussive, interrogative* (Fava, 1995) –, many other new illocutionary types may be envisaged.

Over the past 20 years the LABLITA team has carried out empirical research on spontaneous spoken corpora for the identification of illocutionary types and their prosodic profiles, following a corpus-based methodology supported by experimental tests (Cresti, 2005, 2018; Cresti & Firenzuoli, 1999; Cresti, Moneglia, & Martin, 2003; Firenzuoli, 2003; Rocha, 2016).

The classification task was made even more difficult by the appearance of spoken corpora, since Searle's taxonomy (Searle, 1969) and its extension (Searle & Vanderveken, 1985; Vanderveken, 1990) are unsuitable for application to spoken texts (Cresti, 2017, 2018). Simple evidence of this exists, for instance, in the lack of basic illocutionary types such as *deixis, refusal, recall*, and *reported speech*, which are usually accomplished in spoken exchanges and are frequently found in spontaneous corpora. We cannot elaborate much on the subject here, but the illocutionary types mentioned are the basis of linguistic interaction. The deixis of an object, a person, an event, using the language (pointing), and "rejection" are known in early acquisition to be the first actions reported linguistically. Furthermore, the fact that you call people or animals that are out of sight or that you call their attention if they are far away or inattentive, seems natural. Finally, reporting one's own words or those of others, but also one's own thinking, is a practice found in all the languages of the world. Even if no performative verb can be imagined on which to base the "translation formula" for these linguistic acts, as with Searle's approach, our conception of pragmatics allows us to consider them between illocutionary types. See Section 3.1 for details.

It must be stressed, however, that the overall problem lies in the fact that this logic framework does not provide any operational instruction for research on real data. In spontaneous speech, it is not easy to accurately identify the linguistic stretch accomplishing the illocutionary act. Spontaneous speech analysis requires the selection of reference units for which not just linguistic but also pragmatic relations hold, allowing the discovery of new and unexpected types with respect to a logical conception of language.

Within L-AcT, the reference unit for speech is given as the utterance. Its definition is pragmatic, and its identification is prosodic.[10] The theory specifies that only one Information unit – the Comment – conveys the illocution and this allows us to overcome a fundamental difficulty of empirical research into illocution. It is the prosodic performance, and specifically that of the Comment unit – which is not usually examined or considered in logic and syntactic studies – that is crucial in deciding the real illocutionary values implemented in speech (Cresti et al., 2003; Firenzuoli, 2003; Rocha, 2016).

[10] The L-AcT approach is far from recent tag-set for the annotation of speech acts such as DIT++, and the Dialogue Annotation and Research Tool (DART) by Weisser (2014, 2018), which are based mostly on syntactic and lexical aspects.

In Section 2, we will present stretches of Italian dialogues that are representative of the pragmatic aspects of speech. Specifically, they demonstrate a continuous illocutionary variation, testifying to the richness of illocutionary types, but also the unpredictability of their occurrence. In Section 3, the paper will briefly sketch the corpus-based methodology used for the induction of speech act types from spoken corpora. The open repertory of illocutionary classes, their sub-classes, and the types derived from the work are reported in the Appendix. In Section 4, some notes on the LABLITA description of prosody are discussed. Finally, in Section 5 we take specific examples from our findings and show the conventional pragmatic and prosodic features that allow sharp distinctions between illocutionary types within the assertive sub-classes: within the weak subclass *self-conclusion* versus *assertion taken for granted*, and within the strong subclass *ascertainment* versus *assertion of evidence*.

2. Some examples of spontaneous speech

2.1 A single turn

Let us inspect Example (1) taken from the LABLITA Corpus, which shows a young woman making photocopies for some students and asking a professor if he too needs a copy. The bare transcription of the sequence, which is performed without any pauses, is not easy to interpret nor to segment into its proper reference units (which each accomplishes an illocution):

(1) *lei gliene serve una anch'a lei una in più o no no lei ha questa*
‘you you need one also for you one more or no no you have this one’

However, on listening to the audio and evaluating the f0 tracks (Figure 1) we recognize the performance of four utterances which are demarcated by terminal prosodic breaks (1a).[11] According to the L-AcT system, each utterance accomplishes an independent illocutionary type. They center on the interaction with the addressee and nearly all are illocutionary types (*request of confirmation, ascertainment*) that were not considered in the standard taxonomy (Searle, 1969).[12]

[11] The prosodic parameters correlating with an utterance's performance are analyzed using the software WINPITCH.

[12] See Section 5.4.

(1) a. *SUS: *lei /*^{TOP} *gliene serve una anch'a lei ?*^{COM} *una in più / o no ?*^{COM} *no //* ^{COM} *lei ha questa //* ^{COM}

'you /(do) you need one also for you? one more / or not? no // you have this one //'[13]

%ill[14]: [1] request of confirmation; [2] alternative question; [3] answer; [4] ascertainment

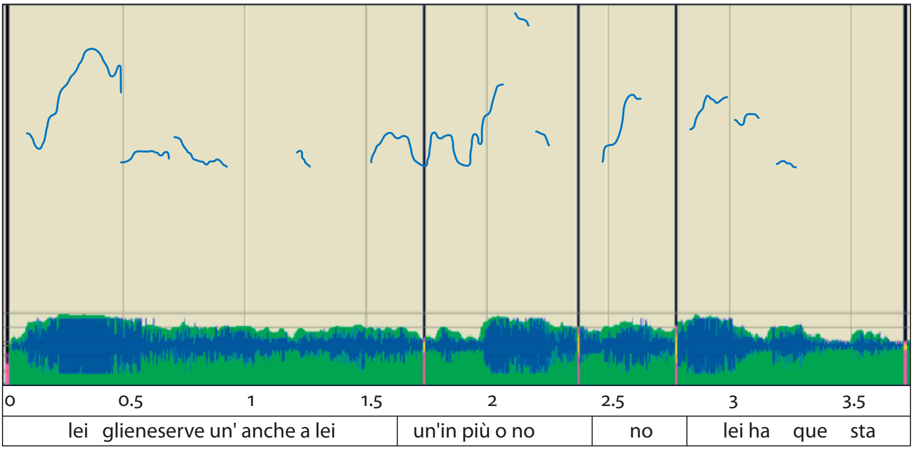


Figure 1. F0 track for Example (1a)

It should be noted that the occurrence of each illocutionary type is unpredictable since it is accomplished through a quick change in SUS's mental representations and pragmatic activity as she relates both to the behavior of the participants and the context. The girl passes from a request for confirmation – which depends on her false hypothesis that the professor needs a photocopy – to an alternative question, to a negative self-answer, to a final assertion stemming from the observation that the professor already has one.

[13] The transcription of spoken texts is in the LABLITA format (Moneglia & Cresti, 1997), which is derived from the CHAT system (MacWhinney, 2000). Each slash gives its information tag using three capital letters in superscript. So far, the corpus-driven classification of information types covers Textual functions, encompassing the Comment (COM), Topic (TOP), Appendix of Comment (APC), Appendix of Topic (APT), Parenthesis (PAR) and Locutive Introducer (INT), and Dialogical functions, encompassing the Incipit (INP), Phatic (PHA), Allocutive (ALL), Conative (CNT), Expressive (EXP) and Dialogical Connector (DCT).

[14] Illocutions.

2.2 A negotiation

Let us look at Example (2), which is an excerpt from a dialogue in a motor shop between a seller and a woman who wants to place an order for a Vespa as a Christmas present for her daughter.

(2) *ALE: *sera* //$^{\text{COM}}$ *arrivo* /$^{\text{COM}}$ *eh* //$^{\text{PHA}}$ *finisco di mettere* /$^{\text{SCA}}$*# in esposizione <i veicoli>* //$^{\text{COM}}$
‘(good) evening // I’m coming / eh // I’m finishing putting these vehicles on display //’
%ill: [1] welcome; [2] waiting request; [3] explanation
%ref.un[15]: [1] utterance; [2] utterance; [3] utterance
*GAB: *<faccia>* /$^{\text{CMM}}$ *faccia pure* //$^{\text{CMM}}$ *perché tanto ho tempo* //$^{\text{COM}}$ *non c’è premura* //$^{\text{COM}}$ *sabato sera yyyy…*$^{\text{COM}}$
‘let’s go / let’s go // because I have time // there’s no hurry // Saturday evening…’
%ill: [1] agreement; [2] softening; [3] softening; [4] expression of obviousness
%ref.un: [1] pattern (reinforcement); [2] utterance; [3] utterance; [4] utterance
*ALE: *yyyy # ecco qua* //$^{\text{COM}}$ *mettiamo a posto <questi>* //$^{\text{COM}}$
‘here I am // I’m putting in place these //’
%ill: [1] conclusion; [2] on-going comment
%ref.un: [1] utterance; [2] utterance (overlapped)
*GAB: *<io cer>* /$^{\text{SCA}}$ *cercavo una vespa* //$^{\text{COM}}$
‘I was / was looking for a Vespa //’
%ill: [1] assertion taken for granted
%ref.un: [1] utterance
*ALE: *sì* //$^{\text{COM}}$
‘yea //’
%ill: [1] assent (dialogical move)
%ref.un: utterance
*GAB: *cinquanta* //$^{\text{COM}}$ *non so se usata* /$^{\text{CMM}}$ *nuova* /$^{\text{CMM}}$ *ha qualche cosa* ?$^{\text{CMM}}$
‘fifty // I don’t know if second-hand / new / have you anything?’
%ill: [1] instruction; [2] yes-no question
%ref.un: pattern (list)
*ALE: *guardi* /$^{\text{CNT}}$ *&he* /$^{\text{TMT}}$ *in questo momento* +$^{\text{TOP}}$ *beh* /$^{\text{PHA}}$ *se la vuole nuova* /$^{\text{TOP}}$ *c’è una bella promozione* /$^{\text{COM}}$ *che abbiamo* /$^{\text{SCA}}$ *<adesso>* //$^{\text{APC}}$
‘listen / uh / at the moment + well / if you want a new one / there is a good promotion / that we have / now //’

[15] Reference unit.

%ill: [1] interrupted; [2] ascertainment
%ref.un: [1] interrupted; [2] utterance
*GAB: *mah* /$^{\text{EXP}}$ *costa un <po' cara>* /$^{\text{COM}}$ *nuova* /$^{\text{PAR}}$ *però* //$^{\text{APC}}$
'well / it costs a little too much / new / indeed //'
%ill: [1] contrast
%ref.un: [1] utterance [source: pubdl11]

The example corresponds to eight turns and is composed of at least 15 reference units plus an interrupted unit. All of the reference units in (2) are utterances, but two of them contain a common spoken strategy: the *illocutionary pattern*. An utterance may correspond to a chain of two or more Comments which can double (or repeat even three times) the same illocutionary act, changing its content or – less frequently – maintaining the same linguistic content. There are many illocutionary patterns which are conceived according to a kind of natural rhetoric model and form a chain of rhythmed multiple Comments (CMM). The most common of these is the *reinforcement* pattern, composed of a doubling of the illocution; alternatively, there are *comparison* and *alternation* patterns, as well as a chain of three or more Comments which make up a *list*.[16] The first illocutionary pattern in (2) represents the doubling of an *agreement* and the other a list of *questions*.

The development of the dialogue is grounded in the continuous variation of the illocutionary forces and their accomplishment is impossible to predict prior to their occurrence as it depends on the free initiative of the speakers.

2.3 An interactive multi-dialogue

Let us look at (3), an instance of a highly interactive multi-dialogue between the owner of a house and two workers who are repairing the gutters and the chimney on the roof. It corresponds to four turn-taking and is composed of eight utterances with mostly directive illocutions (*order, polar question*). The pragmatic goals underlying the interactions between the speakers motivate different illocutions, although all illocutionary types center around the same subject (*to fix something through screws*).

(3) *OLV: *Marco* /$^{\text{CMM}}$ *vieni qui a mettere i fili* //$^{\text{CMM}}$ *dai* //$^{\text{COM}}$
'Marco / come here and fix the wires // do it //'
%ill: [1] recall + order; [2] order
%ref.un: [1] pattern (functional recalling); [2] utterance
*MAR: <xxx> *vai* /$^{\text{CNT}}$ *viti* //$^{\text{COM}}$ # *ce l'hai il cacciavite* ?$^{\text{COM}}$
'O.K. / screws // do you have the screwsdriver?'

[16] According to the IPIC DB, in 21,007 terminated sequences 7.80% are Illocutionary Patterns.

%ill: [1] order; [2] polar question
%ref.un: [1] utterance; [2] utterance
*LEO: *ce n'ha punti qui* ?$^{\text{COM}}$ *sì* /$^{\text{PHA}}$ *ce l'ho* / *ma a stella* //$^{\text{COM}}$
'has he one here? Yea / I've got it / but (it's) a star (screwdriver) //'
%ill: self-question; answer
%ref.un: [1] utterance; [2] utterance
*MAR: *vai* //$^{\text{COM}}$ *va bene questo* /$^{\text{COM}}$ *vai* //$^{\text{PHA}}$
'go on // this is O. K. / go //'
%ill: [1] agreement; [2] confirmation
%ref.un: [1] utterance; [2] utterance [source: pubcv26]

Some illocutionary types illustrated in (3) are new with respect to the Searlian taxonomy (self-question, agreement, confirmation). However, their occurrence appears being still not predictable (e.g., answering an *order* with another *order;* answering a question with a *self-question).* It must be noted that also an illocutionary pattern is performed here, called *functional recalling*. It is composed of a recall illocution followed by an order that corresponds to a kind of natural pragmatic model.

2.4 A family conversation

Let us look at (4), a stretch of conversation between three speakers who are at home calmly looking at old family pictures.

(4) *ELA: *o chi l'è questa* ?$^{\text{COM}}$
'who is that one?'
%ill: (partial question)
*LIA: *'un c' indovini* // $^{\text{COM}}$
'(you) cannot guess //'
%ill: expression of challenge
*MAX: *no* // $^{\text{COM}}$ *'un ci credo* / $^{\text{COM}}$ *no no* //$^{\text{PHA}}$ *ma tu se' te*?$^{\text{COM}}$
'no // I can't believe it / no no // But it is really you?'
%ill: [1] disconfirmation; [2] expression of disappointment; [3] request of confirmation
*LIA: <*no*> // $^{\text{COM}}$
'no, (it is not me) //'
%ill: disconfirmation
*ELA: <*no*> // $^{\text{COM}}$
'no, (it is not she) //'
%ill: assertion (confirmation)
*MAX: *chi è* / *Sonia* ? $^{\text{COM}}$
'it isn't / Sonia ?'

%ill: direction (request of confirmation)
*LIA: *è la Malvina* // COM
'here is Malvina //'
%ill: presentation
*MAX: *mamma <mia>* // COM
'my mother //'
%ill: expression (expression of disappointment) [source: famcv01]

Given the family situation, we might foresee a lower action involvement, but the speakers perform eight quick turn-taking accomplishing ten utterances, with a high rate of illocutionary variation: Eight of the 10 illocutionary types differ from one another and belong not only to the assertive and expressive classes but to the directive class as well. None of them could have been previously foreseen, as for instance the answer to a question with an expression of challenge.

In the previous four examples a continuous illocutionary variation is enacted by the speakers. It should be noted that this is not dependent on the speaker's being within a dialogue or a conversation, being among family or at the University or in the working place. We could go on presenting (many) examples of different pragmatic situations and languages – ranging as widely as English, French et even Chinese and Japanese[17] – but the only thing shared by all of the stretches is their aspect of interactive and spontaneous exchange; indeed, a monologic text would be quite different with respect to its pragmatic aspect and, as a consequence, its illocutionary variation (Cresti, 2019).

In conclusion, it is possible to systematically analyze spontaneous speech through the prosodic identification of reference units and their correlating pragmatic aspect. Furthermore, corpora demonstrate on the one hand the richness of illocutionary types, which cannot conceivably be reduced to fit the traditional taxonomies, and on the other the difficulty, or better the impossibility, of predicting what the next illocution will be.[18]

[17] See Cresti & Fujimura, 2018; Cresti & Moneglia, 2018; Cresti & Moneglia, this volume, Part II; Cresti, Moneglia, & Tucci, 2011.

[18] Under this regard, L-AcT diverges from a framework such as Conversation Analysis (Sacks, Schegloff, & Jefferson, 1974; Schegloff, 1986, 2007), which has tried to approach dialogue development with a game model strategy of few, highly predictable moves (Carletta et al., 1997; Carlson, 1983; Reed, 2006)

3. The L-AcT classification system

3.1 Background

As we have anticipated, the theoretical background which led to the L-AcT classification system for illocution assumes that pragmatics does not coincide with the speaker's ability to correlate speech with context (Krifka, 2007), but rather centers around the speaker's linguistic interaction with the addressee. L-AcT provides an organic explanation of speech from the perspective of the speaker performing the speech act and takes into consideration the deeply-rooted foundations of speech, which originate in mental representation and the affective relationship with the addressee (Fagioli, 2010). We have not the space in this work to further explore this topic (Cresti, 2005, 2017, 2018), we only summarize here that the basis of speech activity is the affect the speaker intends toward the addressee. Specifically, the affect that originates in the perlocutionary activity becomes in the illocutionary activation a *specific action schema*. Speech depends, in effect, on the libidinal asset of the speaker and is characterized by different qualities and degrees of activation. In accordance with the Human Birth Theory, these can be traced back to a human interest in the addressee or to levels of negation or even annulment of his human essence.

The different affective activations, which can be appreciated in observing the corpus, have led us to identify some illocutionary classes. They are based both on the speaker's libidinal asset and on the resulting relationship types established with the addressee: *refusal, assertion, direction, expression*, and *ritual* (Cresti, 2017, 2018).

Superficial observations may misguide and lead one to believe that the illocutionary classes are the same as those in Searle's taxonomy, given that the latter corresponds to a set of five classes also (*representative, directive, commissive, expressive, declarations*). These classes could superficially be confused with the L-AcT repertory because, number aside, they also partly share terminology (*assertion, direction, expression*) and because L-AcT's Ritual class, which is not present in the Searle tradition, could be compared with *declaration*. However, L-AcT adds a new class, the Refusal, and lacks the commissive class. Therefore, the L-AcT proposal for illocutionary classification must be considered as different, both from a terminological point of view and, evidently, with respect to its *substance* (the real difference).

L-AcT departs strongly from Searle's classification (Searle, 1969) since the latter is based on an *effability* principle (Katz, 1977) that equates a performative proposition with an utterance accomplishing an illocutionary act (*I ask you what time it is = what time is it*?). From L-AcT's perspective the two entities present not just differences in style between two propositions which render the illocution fully explicit (or implicitly express it in some way), but, in fact, yield substantially different illocutionary values. In the Searlian point of view the pragmatic essence of the speech act is reduced to only a *pragmatic meaning* that belongs to the locutive act; consequently, the pragmatic specificity of the illocution is unaddressed and disappears. In contrast, L-AcT is based on the pragmatic essence of the illocution and demonstrates its importance and specificity, since it makes a radical distinction between the illocutionary act and the locutionary act. One might think that at least the assertive class could coincide with Searle's, given that a kind of "collapse" between the locutionary and illocutionary levels might be imagined. This would not be insignificant since a little over half of illocutions belong to assertive types, even in spontaneous speech corpora. However, it will be shown in subsequent paragraphs that the assertive class cannot be reduced directly to the locutive fulfilment. The empirical research carried out on corpora, has indeed identified a wide variation of pragmatic traits affecting the accomplishment of assertive types and has led to a rich classification comprehending two illocutionary sub-classes and many illocutionary types, that cannot be expressed and retraced to their locutive fulfilment.

L-AcT also differs from some authoritative proposals that assume that only the change and transformation of the world – following the statement of an utterance – ensures that an illocution has been accomplished (Sbisà, 1989): Only the effect or the set of effects achieved in the world by the utterance, which is recognized legally or by convention, should guarantee that a certain illocutionary force occurred. Thus, even the classification of an illocutionary type should be defined only *a posteriori* by the effects that it causes in the context. From the L-AcT perspective, the illocutionary activation (originating from the affect) is accomplished regardless of its subsequent recognition and takes place in the world even in the absence of acceptance or understanding by some party.

3.2 Pragmatic features

Within L-AcT each illocutionary class extends to a set of illocutionary types which have been discovered through empirical research on corpora. To be used within a social community with the aim of being recognized and understood, the mental representation (intentionally directed by an affect toward an addressee) must be conventionally codified as an illocutionary type. Thus, although action schemas underlying each illocutionary type must be traceable to an affective intention, they must also be transformed into a conventional type and they are classifiable by way of their pragmatic features. The accomplishment of an illocutionary type is driven by a conventional pragmatic *form* which is shared by the speakers of the language community and may in fact be common to an even wider cultural community that extends beyond linguistic boundaries.

At this time, it is not possible to explain the pragmatic features in detail. However, the frequent recurrence of types in Romance corpora as well as in English and as Chinese and Japanese (which early research seems to confirm, see Cresti & Fujimura, 2018; Cresti & Moneglia, 2018) allows us to at least propose them as a reasonable basis for study. As is the case in empirical research, the list of features characterizing illocutionary types may not be exhaustive since the analysis process means the discovery of new aspects and the correction and further detailing of those already identified (they range from communicative, to perceptual, cognitive, intentional, social, and linguistic domains, and they may or may not be present in a type or may participate with varying relevance and to different degrees in others). Still, their enumeration may offer an idea of the domains we have considered so far in this investigation. The features we have commonly used to classify illocutionary types are reported in Table 1.

Table 1. Types of illocutionary features

Feature types	**Features**
Communication	Channel Attentional horizon Focus Context Reference object
Proxemics	Space relations between participants and their movements Gesticulation Gaze
Social	Speaker roles and conditions Addressee roles and conditions
Speaker activity	Intentional values Speaker commitment to the truth Speaker affective involvement
Expected effects	Conventionally expected effects on the addressee Conventionally expected effects in the context Fulfilment time Benefit
Linguistic	Locutive performance Voice and phonetic performance Prosody

Each illocutionary type is identified in the corpus through the clustering of pragmatic features and via its prosodic performance.

3.3 Working procedure for the identification of an illocutionary type in the corpus

The working procedure for the identification of illocutionary types in corpora may be summarized as follows:

1. Collection of at least 10 utterances judged by researchers to convey the same illocutionary type;
2. The type concerned is described in a detailed manner by each researcher, who considers all domains of pragmatic features in extracting basic pragmatic characteristics for the type;
3. A comparison and choice between the features identified by the researchers allow for the outline of a working description of the illocutionary type, focusing on basic aspects and conditions;
4. Following the working description, a script is produced for structuring an artificial situation in the laboratory with the aim of eliciting corresponding illocutions;
5. From the analysis of the linguistic contents of corpus instances some utterances are composed as prototypical examples of the illocutionary type;
6. Actors are asked to play out the prototypical examples in the eliciting situation while being filmed and short scenes are produced;
7. The scenes are verified to evaluate if the actors' performances are suitable in terms of pragmatic naturalness and to see if their prosodic performances are at least comparable with those of the corpus instances;
8. A process for the adjustment and correction of the script, the eliciting situation, and the linguistic characteristics of the prototypical examples goes on until the researchers are satisfied that the filmed scenes present the illocutionary type correctly;
9. A process of testing and group validation starts, concerning judgment of the suitability of the representations of the illocutionary type. The validation looks at the recognition, interpretation, and evaluation of the pragmatic value of the scenes;
10. A separate validation group is set up for the judgment of the prosodic profile's appropriateness and for the development of difference proofs afterward.

The validation process allows the extraction of proper pragmatic features for the illocutionary type, ending with a working description and a preliminary identification of the type. For instance, we may discover that the possibility of performing a self-conclusive illocutionary type depends on whether the speaker ceases to look at his addressee and looks down. Using the same words, an actor can perform an imperative or instructive illocutionary type, depending on proxemics, gaze and overall physical attitude towards the addressee, as well as on the speed of execution (explained in Section 4.2). The same may be noted with regard to the illocutionary types of obviousness and of expression softening, where the first involves a raised gaze and the second a bending of the head on one side and accompanying hands.

4. Prosodic research

4.1 Description of the prosodic nucleus

Meanwhile, a parallel investigation into prosodic profiles is carried out. As anticipated, only the Comment information unit conveys the illocutionary value of the utterance and is performed using a dedicated prosodic unit of the root type, according to IPO terminology. It is important to note that root units have different formal variations that correlate with the expression of specific illocutionary values.

According to LABLITA research, the root unit may be composite, allowing: (1) *preparation*, (2) *nucleus*, (3) *tail*. But a root unit can be also simple, since the nucleus is the only necessary component which determines the variance and it strictly correlates with the expression of the illocutionary force (Cresti, 2011). Our description of the nucleus deals with all prosodic parameters, taking into account intensity, syllabic length, speed and phonetic accuracy, which are relevant in distinguishing prosodic root types from one another.

The f0 movements of the nucleus are of course crucial, but the High/Low characterization of pitch on its own, typical for instance in the Auto-segmental Model (Goldsmith, 1990; Pierrehumbert & Hirschberg, 1990), in our opinion is not sufficient. Other features are considered, as the form of the movements (*rising, falling, platform*), their possible composition, the levels they reach between the start and end points (*low, mid, high, very high*) as a kind of gradation, the modality of the movement (*rapid, slow*), the length of the movement (*short, long*), the timing in the syllable, as well as other execution aspects.

It should be underlined that all values concerning the levels of the movements, their speed and their duration are relative and not absolute quantitative data that can be numerically measured and consequently classified. They depend on the speaker's gender, on his mood and education, on the dia-phasic characteristics of the exchange, on the type of text… It's only in taking these features into consideration that prosodic root types can be recognized. They are clearly distinguishable by native speakers, since they correlate with the expression of illocutionary types (order vs. instruction; discussed in 4.2).[19]

[19] See Section 5 for the distinction between the prosodic nuclei of assertive illocutionary types.

4.2 The distinction between order and instruction

On this point, we'd like to further detail the LABLITA system of prosodic analysis. Taking an example, let us look at the difference in prosodic profile between the root units conveying an order illocution in Example (5) and Figure 2, and an instruction illocution in Example (6) and Figure 3.[20]

(5) *MAX: *ferma* // COM
'stop'
%ill: order [source: famdl13]

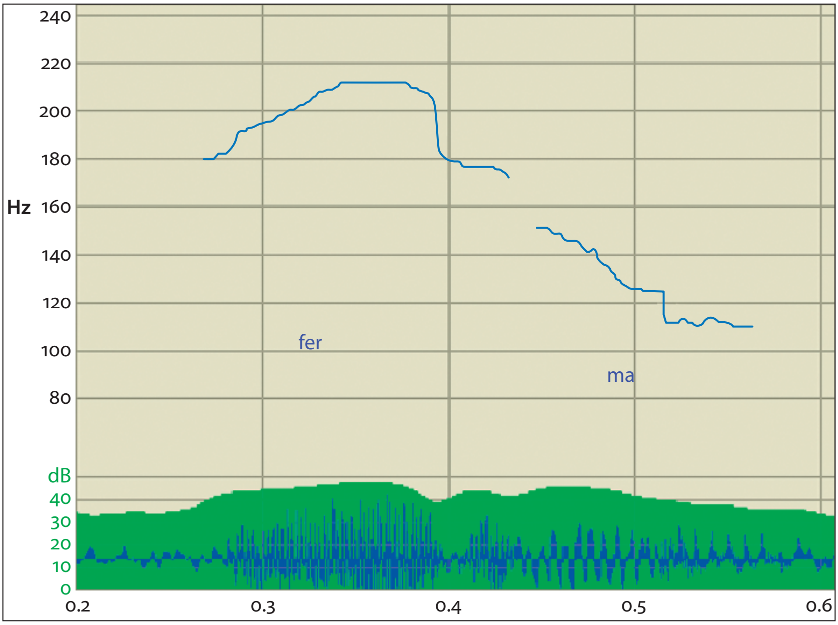


Figure 2. F0 track of Example (5)

(6) *OLV: *tieni con due mani* // COM
'hold (it) with both hands'
%ill: instruction [source: pubcv26]

[20] For the distinction between order and instruction – both pragmatically and prosodically – in Brazilian Portuguese, see Rocha (2016).

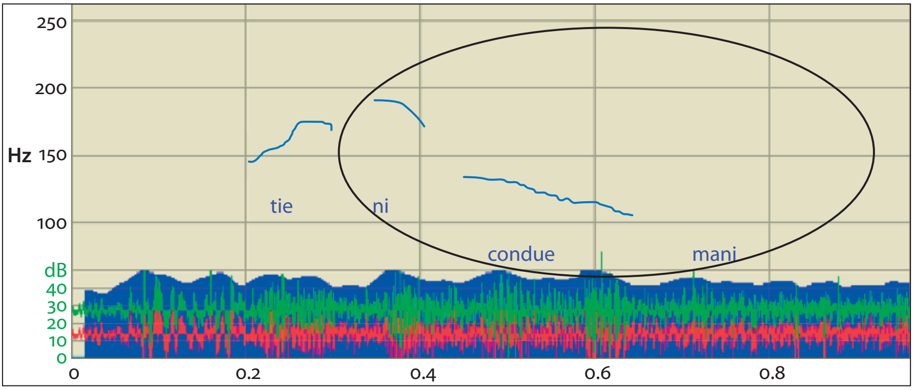


Figure 3. F0 track of Example (6). The nucleus of root unit corresponds to the syllables *ni-con-due-ma-ni*, while the syllable *tie* corresponds to its preparation

The order and instruction illocutionary types are objectively distinguishable through their pragmatic and cognitive features, which pertain to two directive subclasses. While order can be paraphrased as a request to the addressee of intervention in the world, instruction is rather a request to the addressee of his own mental transformation.

Their root profiles may also be distinguished, but given that they in some sense are similar, is possible to get this result if they are analyzed using the right set of features. They are listed in Table 2.

Table 2. Prosodic features of the nucleus for the root unit in an order and an instruction

	Order	**Instruction**
Root composition	+/– preparation, +/– tail	+/– preparation
Nucleus	compound	simple
Form	[Rising-rapid; short; (start High- top very High)] + [Falling-rapid; short; (start High- end Low)]	[Falling-slow; long; (start Mid/ High – end Low)]
Timing	[Rising-rapid + Falling rapid] occur in the same tonic syllable	The continuous Falling movement is spread on all the syllables.
Speed	high	mid/slow
Intensity	strong	mid
Phonetic accuracy	mid	accurate

Studies dedicated to explicating the relationship between root units and illocutionary forces, as well as the conditions governing the performance of the various illocutionary act types, have been carried out for Italian (Cresti, 2005, 2018, forthcoming a; Cresti & Firenzuoli, 1999; Cresti et al., 2003; Firenzuoli, 2003; Moneglia, 2011) and for Brazilian Portuguese (Rocha, 2016).

Returning to the procedure for the identification of an illocutionary type in the corpus, the prosodic profiles of root units – extracted from the corpus examples and evaluated as conveying a specific illocutionary type – are verified and checked to see if they are *compatible* with one another. They are then described systematically, taking the set of appropriate parameters into account. Actors are requested to perform in the validated eliciting situation (after point 8) for some verified prototypical utterances of a certain illocutionary type, and their compliance with the illocutionary type is confirmed. In the final stage, difference proofs are carried out. Laboratory experiments with prosodic normalizations and the synthesis of different prosodic parameters lead to a hypothesis on *the model of a prosodic profile with a specific illocutionary value*.

5. A first classification

5.1 A general overview

The systematic analysis of entire spoken texts allows researchers to recognize the existence of illocutionary types recurring within dia-phasic and dia-stratic variations. The set of pragmatic and prosodic features provides an operative criterion for identifying illocutionary types which are empirically induced.

Our corpus-based research has led to an initial classification of almost 90 illocutionary types which are grouped into five illocutionary classes (*refusal, representation, direction, expression, ritual*) depending on the basic pragmatic/affective activation. In turn, the illocutionary classes can be divided into pragmatic sub-classes which present intermediate levels within each class (see the Appendix to this paper). Therefore, beyond originating in a basic affective activation, each type belongs to a sub-class which shares a cluster of pragmatic features. This repertory is a working set and is open to the addition of new entries which may be discovered during further corpus-based investigations.

Although each illocutionary type belongs to a sub-class together with other illocutionary types, each one must be clearly distinguishable from the other by way of some idiosyncratic features and an individual prosodic profile. The examples reported in the paper give evidence for this claim.

5.2 The assertive class and its sub-classes

Even if the assertion is the most common illocutionary class employed in speech it presents general aspects that, in our view, have not been dealt with in the literature before since they could only be observed via empirical research on corpora.[21]

The assertive class, indeed, is not a monolithic entity and does not correspond to just one single illocutionary type, since there is at least an intermediate level composed of two sub-classes: *weak assertion* and *strong assertion*. The salient features distinguishing the two assertive sub-classes may be summarized as: the degree of relevance of the semantic content in the utterance, the (speaker's) commitment to the content's truth, and the degree of the speaker's involvement with respect to the addressee.

So far, each of the two sub-classes comprehends a variety of different illocutionary types; within the weak-assertion sub-class we see *self-conclusion, on-going comment, confirmation, neutral assertion /explication, assertion taken for granted, literal citation*; and within strong assertion *answer, ascertainment, assertion of evidence, hypothesis.* Most of these assertive types are performed through root units with specific prosodic profiles that convey corresponding forces.

In the paragraphs to follow we provide examples of the *self-conclusion* and the *assertion taken for granted* types, taken from the weak assertion sub-class, and examples of the *ascertainment* and the *assertion of evidence* types, taken from the strong assertion sub-class.

5.3 Examples of the self-conclusion type from the weak assertion sub-class

The utterances with a *self-conclusion* illocution, although performed with commitment to the truth value of their contents, seem rather unconcerned with the addressee's involvement. In these cases, when the speaker is participating in a situation or a conversation with the addressee and arrives at the end of the discussion or mental activity, he seems to suddenly become distant from the flow of the exchange and, without looking at the addressee, begins speaking in a low voice, not caring if the other participant can hear clearly or not. It's as though he were speaking for his own benefit only, even though his assertion is functional with respect to furthering the dialogue.

[21] A reasonable estimate leads us to suppose that for spontaneous interactive speech about 45% of utterances are non-assertive (Firenzuoli, 2000).

Contrary to what might be expected, this *self-conclusion* type has frequently been found in Romance corpora. Let us look at (7) and (8), whose transcriptions in bold underline the specific illocutionary type.

(7) *LIA: ***non si muore*** //COM
'we don't (completely) die //'
(reflecting during a family meeting about the resemblance between the grandson's face and that of her dead husband)
%ill: self-conclusion [source: famcv01]

(8) *LAL: *io 'un son un giocatore* // COM *l'ho detto all'inizio* // COM *sono un giocatore dilettante* // COM ***gioco così per giocare*** //COM
'I'm not a player // I said it at the beginning // I'm a novice player // I play just to play //'
(justifying his loss in a poker game)
%ill: ascertainment; assertion taken for granted; assertion taken for granted; self-conclusion [source: famcv14]

Below are two further examples alongside their prosodic performances. Let us look at Figure 4, f0 track for (9), which is an example taken from the same file as that of (3). LEO is the owner of the house, on whose roof some builders are working. The latter are proposing that he renovate the entire set of tiles. LEO explains to them in a normal tone of voice that he has already had new tiles put down. Then, without looking at them, he says in a low voice that he has no intention of doing this work.

(9) *LEO: *c'è le tegole nuove* / COM *sì* //PHA ***mi fo mett' a fa' un la(v)oro così*** // COM
'here is the new roof tiles / yeah // I can't see myself doing this type of work //'
%ill: assertion (explanation); assertion (self-conclusion) [source: pubcv26]

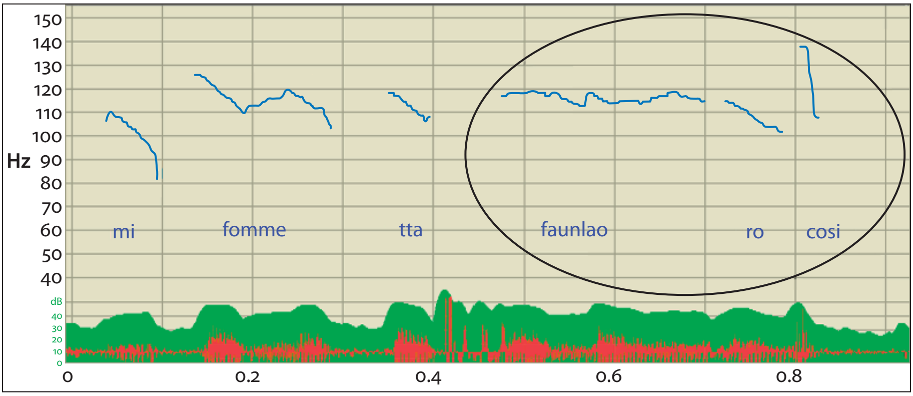


Figure 4. F0 track from Example (9), with stress on the last utterance

Example (10) is taken from the same file as that of (2), where GAB wants to place an order for a Vespa as a Christmas gift to her daughter. The seller, ALE, confirms the order and then with a low voice, as the result of reflection, assures himself that he'll be able to deliver it. Let us look at the f0 track for (10) in Figure 5.

(10) *GAB: *ma per Natale* /TOP *ce la consegnate ancora* /COM <*fosse quella*> ?COM
'but for Christmas / can you still deliver it /<if it were that one>?'
*ALE: <*sì sì sì sì*> //COM ***ce la facciamo ancora*** //COM
'<yes yes yes yes> // there's still time //'
%ill: assertion (confirmation); assertion (self-conclusion) [source: pubdl11]

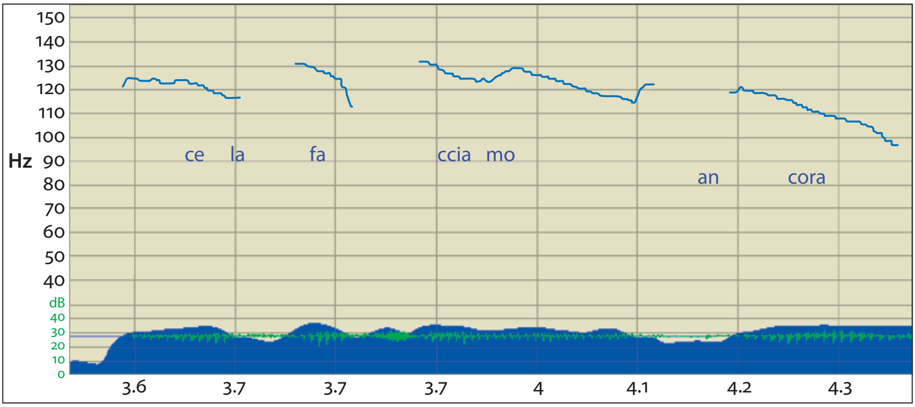


Figure 5. F0 track for Example (10)

Regardless of their differing contents, *self-conclusion* types present the result of speaker reflection and are performed with comparable prosodic profiles. We would like to stress that even if these profiles are specific to self-conclusion, they are mandatorily performed along with a sudden change in interaction with the addressee that implies a lowering of the gaze and a change of voice. The semantic content is not relevant because it doesn't affect the implementation of the illocutionary type. For instance, in (9) the self-conclusion expresses the speaker's intention not to change the roof tiles himself, while in (10) it presents a form of self-reassurance on the part of the speaker that he will be able to deliver the order.

In terms of the LABLITA system, the prosodic profile is described as a simple nucleus that may be preceded by syllables of preparation, but not followed by tail syllables.[22] It corresponds to a unitary falling f0 movement, which is slow and long, beginning with a medium f0 value and ending with a low f0 value. Its intensity is low and the speed rate ranges from medium to slow, while the speaker's voice may be whispered.

[22] Unlike in assertive types, the nucleus of the root units in sub-classes of the directive type (such as *order*, for instance) are often followed by a syllabic tail.

5.4 Examples of the assertion taken for granted type belonging to the weak assertion sub-class

The *assertion taken for granted* type, too, belongs to the weak assertive sub-class and occurs frequently. It occurs within a conversation with one or more addressees, when the speaker tells an old story or reports information that for various reasons can be considered as already known or at least expected at parts. Although the speaker is committed to the truth of the utterance, he doesn't think to offer further insight to the addressees, while presupposing their agreement and awaiting simple acceptance of his report and point of view. We present some examples in (11) and (12).

(11) *ELA: *fino a prima della seconda guerra mondiale* /$^{\mathrm{TOP}}$ ***ci vivevano*** // $^{\mathrm{COM}}$ (as it is well-known)
'until before the second world war / they used to live inside (the Matera's caves)'
%ill: assertion taken for granted [source: famcv17]

(12) *LIA: ***questa è la mi' nonna Stella*** // $^{\mathrm{COM}}$ (everybody knows her)
'this is my grand-mother Stella'
%ill: assertion taken for granted [source: famcv01]

We follow with two more examples, showing their prosodic performances. The context of (13) is the same as that of Example (9) and LEO is recounting his version of a situation in which he made some mistakes, all of which is plain to the workers since they witnessed it. In Figure 6, the f0 for Example (13).

(13) *LEO: *eh* /$^{\mathrm{PHA}}$ *l'avevo infilato dentro un altro tubo* //$^{\mathrm{COM}}$ ***m'è scivolato per la scala*** // $^{\mathrm{COM}}$
'eh / I had shoved it into another tube // it slid down the ladder //'
%ill: assertion taken for granted; assertion taken for granted
[source: pubcv26]

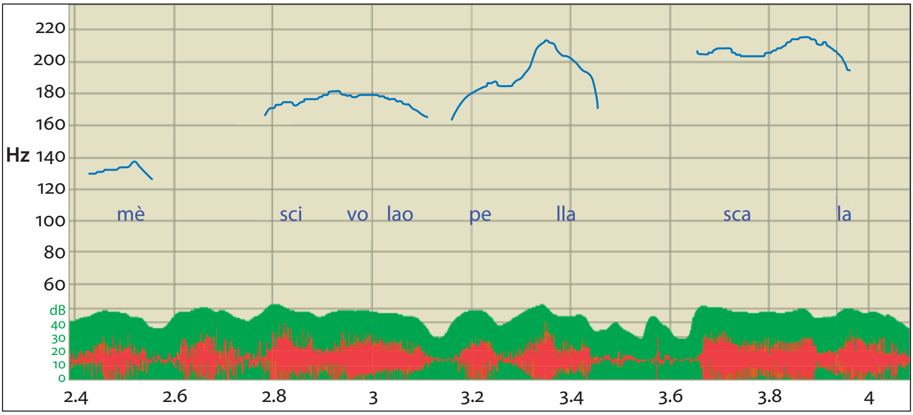


Figure 6. F0 track of the utterance in Example (13)

Similarly, in (14) the situation is the same as in Example (10) and the seller ALE is describing a Vespa model. He begins explaining that it is green and a four-stroke model. At the end of the turn, he repeats that the model is four-stroke, but this time the information has nothing new to impart. In Figure 7, the f0 for Example (14).

(14) *ALE: *questa qui verde /*TOP *è una quattro tempi //* COM *si chiama / verde* [/1] *verde Portovenere //* COM ***è un quattro tempi*** // COM
'this green/ is a four-stroke // its name is / green [/1] green Portovenere // it is a four stroke //'
%ill: assertion (description); assertion (description); assertion (taken for granted) [source: pubdl11]

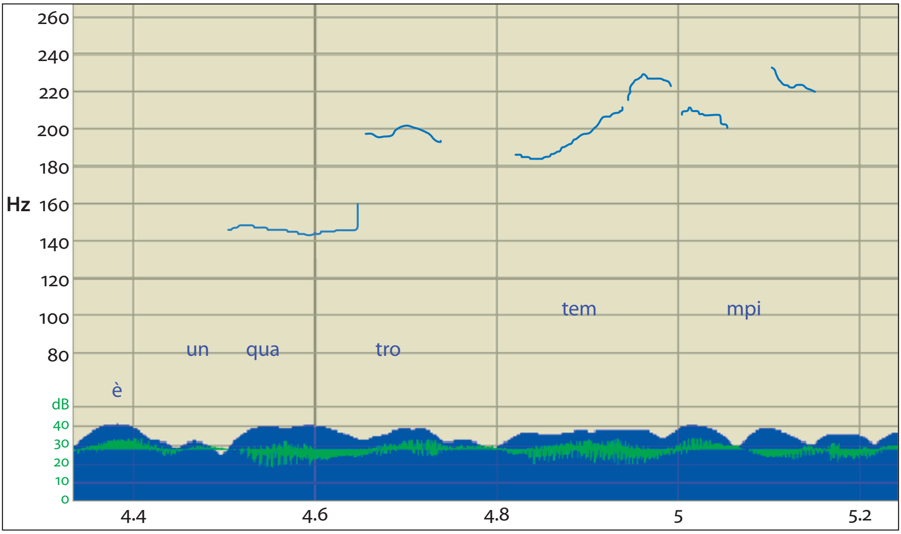


Figure 7. F0 track for Example (14)

Regardless of the differences in their lexical fulfilments, the *assertion taken for granted* types are realized with a root profile whose nucleus may be preceded by preparation syllables and may not be followed by a syllabic tail. The movement of the nucleus is composite and corresponds to a long, ascending prosodic platform, with a medium f0 value followed by a short, final rising movement which terminates in a high f0 value, occurring on the tonic syllable and eventually being lengthened on the post-tonic. The speed rate of the entire profile is quite high.

5.5 Examples of the ascertainment type in the strong assertion sub-class

As we stated previously, the main features which distinguish the strong assertive sub-class from others are the degree of semantic relevance, the speaker's commitment to the truth, and the speaker's degree of involvement with respect to the addressee. Thus, since strongly assertive types are overtly directed at the addressee, they are usually pronounced clearly and with a distinct prosodic profile.

The speaker accomplishes an *ascertainment* illocution because of the observation of a verified state of things. This illocutionary type appears to be concerned with the speaker-addressee exchange, unlike instances of self-conclusion. Utterances with an *ascertainment* illocution are considered relevant to the addressee and therefore the speaker looks at him and speaks in a clear voice, concerning himself with its audibility. Sometimes the type may have a connotation of light disagreement with respect to something mentioned prior by the addressee or happening in the situation. The locutive content may correspond to presentative clauses (*there is*) and to sentences with a deictic subject (*I, this, today*) or a full semantic subject, which are rare in Italian spontaneous speech. Realizing a semantic focus at the beginning of the Comment and marking it with a prosodic prominence, the speaker refers to the *point* he has verified. Some examples are provided in (15) and (16).

(15) *FAB: ***c'è il peperoncino*** // $^{\text{COM}}$
'there is chili (inside) //'
(commenting negatively on the flavour of a dish)
%ill: ascertainment [source: famcv12]

(16) *ILA: ***c'è un'acustica*** / $^{\text{COM}}$ *fa schifo* //$^{\text{APC}}$
'there are acoustics / (that) are awful //'
(stating the poor audio recording quality of the room)
%ill: ascertainment [source: famcv06]

Let us also look at (17) and (18), along with their prosodic performances, respectively in Figure 8 and Figure 9.

(17) *MAX: ***questa è a Londra*** // $^{\text{COM}}$
'this (picture) is (taken) in London //'
(recognizing an old picture)
%ill: ascertainment [source: famcv01]

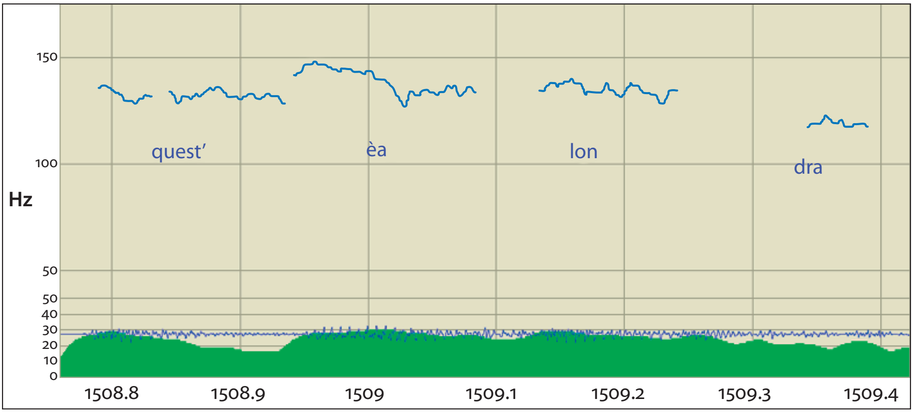


Figure 8. F0 track for Example (17)

(18) *GAL: ***i soldi vanno messi*** // COM
'money must be put on (the table)'
(within a card game because of the lack of money)
%ill: ascertainment [source: famcv14]

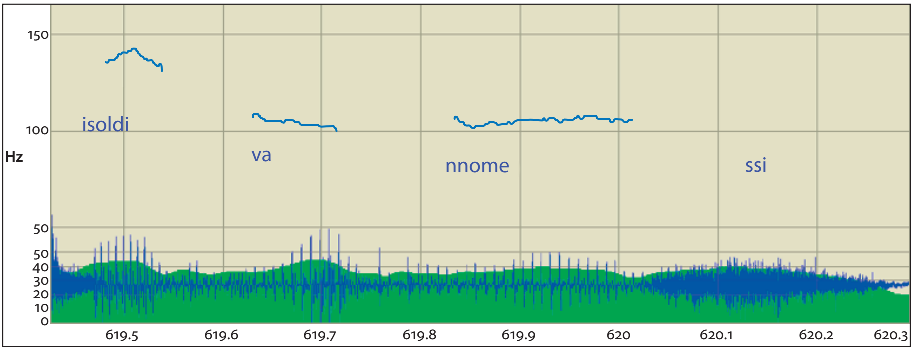


Figure 9. F0 track for Example (18)

Regardless of the differences in lexical fulfilment, the *ascertainment* types are realized with a root profile whose nucleus is composed of a short and rapid rising movement (on the tonic syllable of the first semantic word of the Comment unit) and a long, lengthened movement which descends until it reaches a relatively low level (covering the rest of the Comment).

5.6 Examples of the assertion of evidence type from the strong assertion sub-class

The speaker performs an *assertion of evidence* illocution when attempting to convince the addressee of the "evidence" for his opinion. This act may be compared with the *ascertainment* type, which in some sense is its opposite. With the latter the speaker has a high commitment to the truth of the semantic content being presented because it may be verified instantly in the context, or, in any case has passed through the perception and knowledge of the speaker. In contrast, the evidence type depends on the speaker's conviction of the validity of his opinion, from which derives his intention that the interlocutor agree with him. Thus, the semantic content is a kind of "ethical" evidence for the speaker.

The argument concerning the evidence may be brought into view by the speaker through the employment of a Topic unit. In accomplishing this illocutionary type, the speaker states something and tries to make it evident to the addressee. Unlike in the ascertainment type, the semantic focus occurs at the end of the Comment unit and indicates the evidence on which the addressee should focus his attention. The *assertion of evidence* type, too, appears concerned with the speaker-addressee exchange and is considered relevant to the addressee. Thus, the speaker talks in a clear voice and takes care that he is audible. The type is quite common. Some examples are provided in (19) and (20).

(19) *LAU: *comunque delle pensiline* /$^{\text{TOP}}$ ***le devi creare*** // $^{\text{COM}}$
'however some shelters / you must build them //'
(architect's advice on a project)
%ill: assertion of evidence [source: famcv16]

(20) *VAL: *il marito* /$^{\text{TOP}}$ ***conta poco*** // $^{\text{COM}}$
'the husband / counts little //'
(explaining to a colleague the probable score criterion for getting a job at the school)
%ill: assertion of evidence [source: ifamcv18]

Furthermore, here are (21) and (22), and their prosodic performances, respectively in Figure 10 and Figure 11.

(21) *LAK: ***un' son mica poche*** // $^{\text{COM}}$
'it is not little money //'
(evaluating the pot of poker present on the table)
%ill: assertion of evidence [source: famcv14]

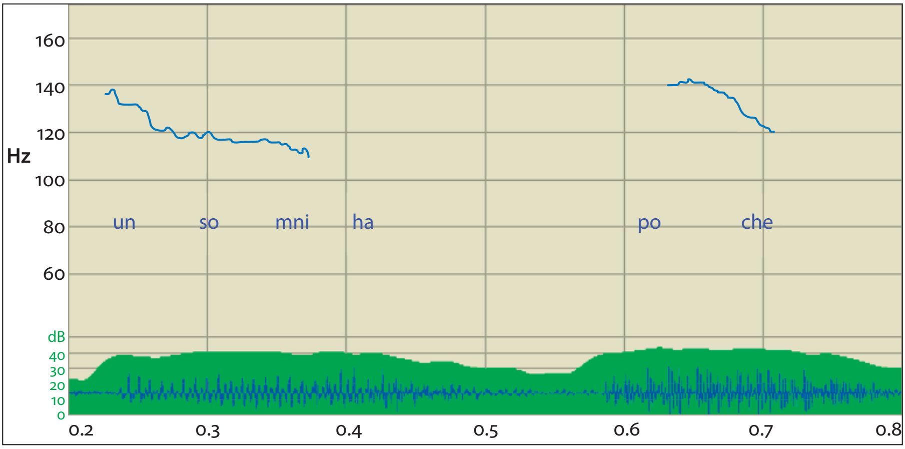


Figure 10. F0 track for Example (21)

(22) *WAL: *ma quando l'hai murate lì* /^{TOP} ***'un importa mica tu metta la staffa*** //^{COM}
'But once you've bricked over it / it doesn't matter if you put a bracket //'
%ill: assertion of evidence [source: pubcv26]

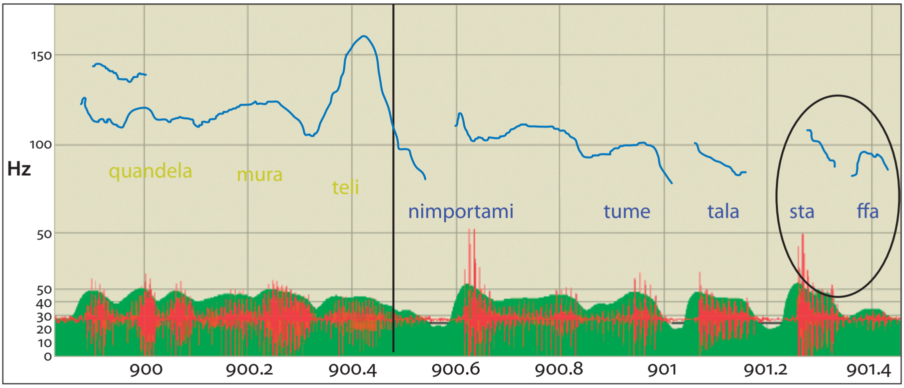


Figure 11. F0 track for Example (22)

Regardless of any difference in lexical fulfilments, the *assertion of evidence* type is realized with a root unit whose nucleus may be preceded by some preparation syllables, but not followed by a syllabic tail. The nucleus is a compound and is made up of a mid-raising movement and a short movement that falls until it reaches mid-level (on the tonic syllable of the last semantic word of the Comment unit). The speed rate of the entire profile is rather high.

6. Conclusions

The primary aim of L-AcT is to ground the systematic analysis of spoken corpora in speech act theory and to this end a corpus-based methodology has been developed. Its main innovation with respect to Austin's work is to consider that the spoken activity (the illocutionary force and the information structure) manifests itself through prosodic devices. Only a corpus providing reference units for speech – that is to say, providing utterances which are text and sound aligned – fully allows an investigation of this type. The LABLITA Corpus with highly-diversified corpus designs of spoken Italian, ensuring an in-depth representation of interactive situations taken from real life, has been collected, analyzed and aligned. The systematic identification of speech reference units has been carried out for it and the IPIC database constructed as a result, presenting information tagging for significant corpus selections. Thus, an empirical investigation on illocution has been carried out for spoken Italian.

L-AcT proposes an accurate system for the illocutionary classification of spontaneous speech. The working procedure has generated an open repertory which is rich and structured in terms of classes, sub-classes, and dozens of illocutionary types which seems to be shared across English and Romance languages. The repertory is a direct result of the L-AcT framework, which considers the speaker's affect and the resulting relationship with the addressee as the foundations of speech. The corpus-driven research, carried out on the Italian corpus, crucially revealed the continuous illocutionary variation and its richness and unpredictability, the explanation of which can be traced to the nature of the human psychic relation.

The assertive class of illocution is the most commonly employed in speech, yet it presents aspects that, in our view, have not been dealt with in the literature up until now, as they could only be observed via empirical research on corpora. Specifically, a sharp distinction between illocutionary types within the assertive class and its sub-classes has become apparent through the discovery of pragmatic and prosodic features that are constant and recurring. According to Italian examples, the assertion repertory comprehends cases such as *self-conclusion* and *assertion taken for granted* within the weak assertive subclass and *assertion of evidence* and *evidentiality assertion* within the strong subclass, none of which, as far as we are aware, have been cited or described before.

References

Arbib, M. (2012). *How the brain got language*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199896684.001.0001

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

Barth-Weingarten, D., Reber, E., & Selting, M. (Eds.). (2010). *Prosody in interaction*. Amsterdam: John Benjamins. https://doi.org/10.1075/sidag.23

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511621024

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (Eds.). (1999). *The longman grammar of spoken and written English*. London: Longman.

Blanche-Benveniste, C. (1997). *Approches de la langue parlée en Français*. Paris: Ophrys.

Blanche-Benveniste, C. (2003). Le recouvrement de la syntaxe et de la macro-syntaxe. In A. Scarano (Ed.), *Macro-syntaxe et pragmatique. L'analyse linguistique de l'oral* (pp. 53–75). Roma: Bulzoni.

Blanche-Benveniste, C., Rouget, C., Bilger, M., & van den Eynde, K. (Eds.). (1990). *Le Français parlé. Études grammaticales*. Paris: Éditions du C.N.R.S.

Buhmann, J., Caspers, J., van Heuven, V. J., Hoekstra, H., Martens, J. P., & Swerts, M. (2002). Annotation of prominent words, prosodic boundaries and segmental lengthening by non-expert transcribers in the Spoken Dutch Corpus. In M. G. Rodriguez, & C. Suarez Araujo (Eds.), *Proceedings of the 3rd LREC conference* (pp. 779–785). Paris: ELRA.

Calesini, I. (2017). Fagioli's human birth theory and the possibility to cure mental illness. *International Journal of Environment and Health*, 8(3), 185–192. Retrieved from https://doi.org/10.1504/IJENVH.2017.086188>

Carletta, J., Isard, S., Doherty-Sneddon, G., Isard, A., Kowtko, J. C., & Anderson, A. H. (1997). The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23(1), 13–31.

Carlson, L. (1983). *Dialogue games: An approach to discourse analysis*. Dordrecht: Reidel.

Cavalcante, F. A., & Ramos, A. C. (2016). The American English spontaneous speech minicorpus. Architecture and comparability. *CHIMERA: Romance Corpora and Linguistic Studies*, 3(2), 99–124.

Chafe, W. (1970). *Meaning and the structure of language*. Chicago, IL: University of Chicago Press.

Chafe, W. (1980). *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.

Cheng, W., Greaves, C., & Warren, M. (2008). *A corpus-driven study of discourse intonation: The Hong Kong corpus of spoken English*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.32

Couper-Kuhlen, E. (2004). Prosody and sequence organizations in English conversation; The case of new beginnings. In E. Couper-Kuhlen, & C. E. Ford (Eds.), *Sound patterns in interaction* (pp. 335–376). Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.62.17cou

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

Cresti, E. (2005). Per una nuova classificazione dell'illocuzione a partire da un corpus di parlato (LABLITA). In E. Burr (Ed.), *Tradizione e innovazione: Il parlato. Atti del VI convegno internazionale SILFI* (pp. 233–246). Pisa: Cesati.

Cresti, E. (2011). The definition of focus in the framework of the Language into Act Theory (L-AcT). In A. Panunzi, T. Raso, & H. Mello (Eds.), *Pragmatics and prosody. Illocution, modality, attitude, information patterning and speech annotation* (pp. 39–82). Firenze: Firenze University Press.

Cresti, E. (2017). The empirical foundation of illocutionary classification. In A. De Meo & F. Dovetto (Eds.), *Atti del convegno, la comunicazione parlata, Napoli, SLI – GSCP International Conference Napoli 2016* (pp. 243–264). Napoli: Aracne.

Cresti, E. (2018). Per una classificazione empirica dell'illocuzione. Lo stato dell'arte. In M. Biffi, F. Cialdini, & R. Setti (Eds.), *"Acciò che'l nostro dire sia ben chiaro". Scritti per Nicoletta Maraschio* (pp. 261–279) Firenze: Accademia della Crusca.

Cresti, E. (2019). Dal polilogo al monologo nell'italiano parlato: La base pragmatico/prosodica del bi-/multi-dialogo e la sua declinazione monologica in testi narrativi e argomentativi. In E. Jamirovsky & M. Durkiewicz (Eds.), *Dal monologo al polilogo: l'Italia nel mondo. Lingue, letterature e culture in contatto*. Kwartalinik Neofilologiczny, 2–2019, vol. III, 341–352.

Cresti, E., & Firenzuoli, V. (1999). Illocution et profils intonatifs de l'italien. *Revue Française de Linguistique Appliquèe*, 4(2), 77–98.

Cresti, E., & Fujimura, I. (2018). The information structure of spontaneous spoken Japanese and Italian in comparison: A pilot study. In A. Manco (Ed.), *Le lingue extraeuropee e l'italiano. Problemi didattici, sociolinguistici, culturali* (pp. 167–189). Milano: Officina 21.

Cresti, E., & Moneglia, M. (Eds.). (2005). *C-ORAL-ROM. Integrated reference corpora for spoken romance languages*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.15

Cresti, E., & Moneglia, M. (2018). The illocutionary basis of information structure. Language into Act Theory (L-AcT). In E. Adamou, K. Haude, & M. Vanhove (Eds.), *Information structure in lesser-described languages: Studies in prosody and syntax* (pp. 359–401). Amsterdam: John Benjamins.

Cresti, E., & Moneglia, M. (this volume). Some notes on the excerpts according to L-AcT. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Cresti, E., Moneglia, M., & Martin, P. (2003). L'intonation des illocutions naturelles répresentatives: Analyse et validation perceptive. In A. Scarano (Ed.), *Macro-syntaxe et pragmatique. L'analyse linguistique de l'oral* (pp. 243–264). Roma: Bulzoni.

Cresti, E., Moneglia, M., & Panunzi, A. (2018). The LABLITA corpus & the Language into Act Theory: Analysis of Viterbo excerpts. In A. De Dominicis (Ed.), *Atti del convegno internazionale "Speech audio archives: Preservation, restoration, annotation, aimed at supporting the linguistic analysis". Accademia Nazionale dei Lincei, CDXV, n.137* (pp. 47–63).

Cresti, E., Moneglia, M., & Tucci, I. (2011). Annotation de l'entretien d'Anita Musso selon la théorie de la langue en acte. *Langue Française*, 2, 95–110. https://doi.org/10.3917/lf.170.0095

Crystal, D. (1975). *The English tone of voice*. London: Arnold.

Danieli, M., Garrido, J. M., Moneglia, M., Panizza, A., Quazza, S., & Swerts, M. (2004). Evaluation of consensus on the annotation of prosodic breaks in the romance corpus of spontaneous speech C-ORAL-ROM. In M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, & R. Silva (Eds.), *Proceedings of the 4th LREC conference* (pp. 1513–1516). Paris: ELRA.

Debaisieux, J.-M. (Ed.). (2013). *Analyses linguistiques sur corpus: Subordination et insubordination en français*. Paris: Hermès-Lavoisier.

Degand, L. & Simon, A. (2009). Mapping prosody and syntax as discourse strategy: How basic discourse units vary across genres. In A. Wichmann, D. Barth-Weingarten, & N. Dehé (Eds.), *Where prosody meets pragmatics: Research at the interface* (pp. 79–105). Bingley: Emerald. https://doi.org/10.1163/9789004253223_005

DIT++ Taxonomy of dialogue acts (Version 5.2) [Computer software]. Retrieved from <http://dit.uvt.nl/>

Du Bois, J. W., Cumming, S., Schuetze-Coburn, S., & Paolino, D. (Eds.). (1992). *Discourse transcription. Santa Barbara papers in linguistics 4*. Santa Barbara, CA: University of California Press.

Du Bois, J. W., Chafe, W. L., Meyer, C., & Thompson, S. A. (2000). *Santa Barbara corpus of spoken American English, Part 1*. Philadelphia, PA: Linguistic Data Consortium.

Egorova, N., Shtyrov, Y., & Pulvermüller, F. (2015). Brain basis of communicative actions in language. *NeuroImage*, 125, 857–867. https://doi.org/10.1016/j.neuroimage.2015.10.055

Fagioli, M. (2010). *Istinto di morte e conoscenza*. Roma: L'Asino d'oro.

Fagioli, M. (2011). *La marionetta e il burattino*. Roma: L'Asino d'oro.

Fagioli, M. (2012). *Teoria della nascita e castrazione umana*. Roma: L'Asino d'oro.

Fava, E. (1995). Tipi di atti e tipi di frase. In L. Renzi, G. Salvi, & A. Cardinaletti (Eds.), *Grande grammatica iItaliana di consultazione* (pp. 19–48). Bologna: Il Mulino.

Firenzuoli, V. (2000). Nuovi dati statistici sull'italiano parlato. *Romanische Forschungen*, 13, 213–225.

Firenzuoli, V. (2003). Le forme intonative di valore illocutivo dell' Italiano parlato: Analisi sperimentale di un corpus di parlato spontaneo (LABLITA) (Unpublished doctoral dissertation). Università di Firenze, Italy.

Gatti, M. G., Becucci, E., Fargnoli, F., Fagioli, M., Ådén, U., & Buonocore, G. (2012). Functional maturation of neocortex: A base of viability. *The Journal of Maternal-Fetal & Neonatal Medicine*, 25(1), 101–103. https://doi.org/10.3109/14767058.2012.664351

Giorgini, L., Petrucci, M., Melcarne, R., Raballo, A., Gatti, M., & Gebhardt, E. (forthcoming). A contribute to the psychotherapeutic treatment according to the Human Birth Theory. In *Proceedings of XII world congress of psychiatry, 27–30th September 2018*. Mexico City: Elsevier.

Goldsmith, J. (1990). *Autosegmental and metrical phonology*. Oxford: Blackwell.

't Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual study on intonation. An experimental approach to speech melody*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511627743

Izre'el, S. (2005). Intonation units and the structure of spontaneous spoken language: A view from Hebrew. In C. Auran, R. Bernard, C. Chanet, A. Colass, A. Di Christo, C. Portes, A. Reynier, & M. Vion (Eds.), *Proceedings of the IDP05 international symposium on discourse-prosody interfaces*. Aix-en-Provence: Université de Provence.

Izre'el, S., & Mettouchi, A. (2015). Representation of speech in CorpAfroAs: Transcriptional strategies and prosodic units. In A. Mettouchi, M. Vanhove, & D. Caubet (Eds.), *Corpus-based studies of lesser-described languages: The CorpAfroAs corpus of spoken AfroAsiatic languages* (pp. 13–41). Amsterdam: John Benjamins.

Karcevsky, S. (1931). Sur la phonologie de la phrase. *Travaux du Cercle Linguistique de Prague*, 4, 188–228.

Katz, J. J. (1977). *Propositional structure and illocutionary force: A study of the contribution of sentence meaning to speech acts*. New York, NY: T.Y. Crowell.

Kempson, R. M. (1977). *Semantic theory*. Cambridge: Cambridge University Press.

Krifka, M. (2007). Basic notions of information structure. In C. Fery & M. Krifka (Eds.), *Interdisciplinary studies on information structure* (Vol. 6, pp. 13–56). Potsdam: Universitätsverlag.

Krifka, M., & Musan, R. (Eds.). (2012). *The expression of information structure*. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110261608

Lacheret-Dujour, A., Kahane, S., & Pietrandrea, P. (Eds.). (2018). *Rhapsodie: A prosodic and syntactic tree-bank for spoken French*. Amsterdam: John Benjamins.

Leech, G. (2014). *The pragmatics of politeness*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195341386.001.0001

Maccari, S., Polese, D., Reynaert, M.-L., & Fagioli, F. (2016). Early-life experiences and the development of adult diseases with a focus on mental illness: The human birth. *Neuroscience*, 342, 232–251. https://doi.org/10.1016/j.neuroscience.2016.05.042

Martin, P. (2015). *The structure of spoken language. Intonation in romance*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139566391

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Mello, H. (2014). Methodological issues for spontaneous speech corpora compilation: The case of C-ORAL-BRASIL. In T. Raso & H. Mello, (Eds.), *Spoken corpora and linguistic studies* (pp. 27–68). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61.01mel

Mello, H., Raso, T., Mittmann, M., Vale, H., & Côrtes, P. (2012). Transcrição e segmentação prosódica do corpus C-ORAL-BRASIL: Critérios de implementação e validação. In T. Raso & H. Mello (Eds.), *C-ORAL-BRASIL I: Corpus de referência de português brasileiro falado informal* (pp. 125–176). Belo Horizonte: Editora UFMA.

Mittmann-Malvessi, M., & Raso, T. (2012). The C-ORAL-BRASIL informationally tagged mini-corpus. In H. Mello, A. Panunzi, & T. Raso (Eds.), *Illocution, modality, attitude, information patterning and speech annotation* (pp. 151–183). Firenze: Firenze University Press.

Mollo, G., Pulvermüller, F., & Hauk, O. (2016). Movement priming of EEG/MEG brain responses for action-words characterizes the link between language and action. *Cortex*, 74, 262–276. https://doi.org/10.1016/j.cortex.2015.10.021

Moneglia, M. (2006). Units of analysis of spontaneous speech and speech variation in a cross-linguistic perspective. In Y. Kawaguchi, S. Zaima, & T. Takagaki (Eds.), *Spoken language corpus and linguistics informatics* (pp. 153–179). Amsterdam: John Benjamins. https://doi.org/10.1075/ubli.5.13mon

Moneglia, M. (2011). Spoken corpora and pragmatics. *Revista Brasileira de Lingustica Aplcada*, 11(2), 479–519. https://doi.org/10.1590/S1984-63982011000200009

Moneglia, M., & Cresti, E. (1997). L'intonazione e i criteri di trascrizione del parlato adulto e infantile. In U. Bortolini, & E. Pizzuto (Eds.), *Il progetto CHILDES Italia* (pp. 57–90). Pisa: Del Cerro.

Moneglia, M., & Cresti, E. (2006). C-ORAL-ROM prosodic boundaries for spontaneous speech analysis. In Y. Kawaguchi, S. Zaima, & T. Takagaki (Eds.), *Spoken language corpus and linguistics informatics* (pp. 89–114). Amsterdam: Benjamins. https://doi.org/10.1075/ubli.5.07mon

Moneglia, M., & Cresti, E. (2015). The cross-linguistic comparison of information patterning in spontaneous speech corpora: Data from C-ORAL-ROM ITALIAN and C-ORAL-BRASIL. In S. Klaeger, & B. Thörle (Eds.), *Interactional linguistics: Grammar and interaction in romance languages from a contrasting point of view* (pp. 107–128). Tübingen: Stauffenburg.

Moneglia, M., & Raso, T. (2014). Notes on the Language into Act Theory. In T. Raso & H. Mello (Eds.), *Spoken corpora and linguistics studies* (pp. 468–494). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61.15mon

Moneglia, M., Raso, T., Mittmann-Malvessi, M., & Mello, H. (2010). Challenging the perceptual relevance of prosodic breaks in multilingual spontaneous speech corpora: C-ORAL-BRASIL/C-ORAL-ROM. In *Speech prosody 2010 conference proceedings*. (pp. 1–4). Chicago, IL.

Nicolás Martínez, M. C. (2012). *Cor-DiAL (Corpus oral didáctico anotado lingüísticamente)*. Madrid: Liceus.

Nicolás Martínez, M. C., & Lombán, M. (2018). Mini-corpus del español para DB-IPIC. *CHIMERA*, 5(2).

Panunzi, A., & Gregori, L. (2011). DB-IPIC. AN XML database for the representation of information structure in spoken language. In H. Mello, A. Panunzi, & T. Raso (Eds.), *Pragmatics and prosody. Illocution, modality, attitude, information patterning and speech annotation* (pp. 133–150). Firenze: Firenze University Press.

Panunzi, A., & Mittmann-Malvessi, M. (2014). The IPIC resource and a cross-linguistic analysis of information structure in Italian and Brazilian Portuguese. In T. Raso & H. Mello (Eds.), *Spoken corpora and Linguistic Studies* (pp. 129–151). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61.05pan

Pierrehumbert, J., & Hirschberg, G. (1990). Intonational phrasing and discourse segmentation. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in Communication* (pp. 271–311). Cambridge, MA: The MIT Press.

Raso, T. (2014). Prosodic constraints for discourse markers. In T. Raso & H. Mello (Eds.), *Spoken corpora and linguistic studies* (pp. 411–467). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61.14ras

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

Raso, T., & Mittmann-Malvessi, M. (2009). Validação estatística dos critérios de segmentação da fala espontânea no corpus C-ORAL-BRASIL. *Revista de Estudos da Linguagem*, 17(2), 73–91. https://doi.org/10.17851/2237-2083.17.2.73-91

Reed, C. (2006). Representing dialogic argumentation. *Knowledge-Based Systems*, 19, 22–31. https://doi.org/10.1016/j.knosys.2005.08.002

Rocha, B. (2016). Uma metodologia empírica para a identificação e descrição de ilocuções e a sua aplicação para o estudo da ordem em PB e Italiano (Unpublished doctoral dissertation). Federal University of Minas Gerais, Brazil).

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation, *Language*, 50(4), 696–735. https://doi.org/10.1353/lan.1974.0010

Sbisà, M. (1989). *Linguaggio, ragione, interazione: Per una teoria pragmatica degli atti linguistici*. Bologna: Il Mulino.

Sbisà, M., & Turner, K. (Eds.). (2013). *Pragmatics of speech actions*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110214383

Schegloff, E. A. (1986). Turn organization: One intersection of grammar and interaction. In E. Ochs, E. A. Schegloff, & S. Thompson (Eds.), *Interaction and grammar* (pp. 52–133). Cambridge: Cambridge University Press.

Schegloff, E. A. (2007). *Sequence organization in interaction. A primer in conversation analysis*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511791208

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139173438

Searle, J. R., & Vanderveken, D. (1985). *Foundations of illocutionary logic*. Cambridge: Cambridge University Press.

Stalnaker, R. (1999). *Context and content*. Oxford: Oxford University Press. https://doi.org/10.1093/0198237073.001.0001

Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *Journal of the Acoustical Society of America*, 101, 514–521. https://doi.org/10.1121/1.418114

Swerts, M., & Geluykens, R. (1993). The prosody of information units in spontaneous monologues. *Phonetica*, 50, 189–196. https://doi.org/10.1159/000261939

Vanderveken, D. (1990). *Meaning and speech acts: Volume 1, principles of language use*. Cambridge: Cambridge University Press.

Weisser, M. (2014). The dialogue annotation and research tool (DART) (Version 1.0) [Computer software]. Retrieved from <martinweisser.org/ling.softhtml#DART>

Weisser, M. (2018). *How to do corpus pragmatics on pragmatically annotated data: Speech acts and beyond*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.84

Appendix. Illocutionary classes, sub-classes and types

Assertion		Direction		Expression		Rite		Refusal
weak	Self-conclusion On-going comment Confirmation Neutral assertion / explanation Assertion taken for granted Literal citation	appearance communicative involvement	Distal recall (non-visible addressee) Distal recall (visible addressee) Proximal recall Functional recall (cmm)	belief	Contrast Softening Expression of obviousness Irony Disbelief /doubt Admission Waiver / renouncement Rhetorical question	courtesy rites (social field: education and civic life)	Thanks Greetings Welcome Excuses Wishes Congratulations Condolences Compliments	
		change of attention	Distal deixis (moving object) Distal deixis (still object) Proximal deixis Prompt Event presentation Mental deixis					
strong	Answer Ascertainment Assertion of evidence Hypothesis	mental transformation	Instruction Introduction of person Request of agreement Self-correction Reported speech Notification /warning	feelings moods state of mind	Protest Complain Grumbling Imprecation Surprise/wonder Wish /desire Easing	bond rites (social field: low, religion, institutions)	Legal declarations Convictions Judgments Penalties Results of examination Medical diagnoses Dedications Religious rites	

Assertion	Direction		Expression		Rite		Refusal
	linguistic behavior	Partial question (information) Polar question (information/behavior) Alternative question (information/behavior) Focalized question (information/behavior) Request of confirmation/ agreement (information/behavior)	speaker / addressee relation	Approval/disapproval Derision Provocation/ challenge Reproach Allusion /hint / negative suggestion Allowance / concession Encouragement / support	dialogic moves	Assent Request of repetition Request to stop Request to wait	
	behavior	Order Interdiction Prohibition Invite Offer Agreement					
	endorsement	Commitment (bet, promise) Proposal Authorization					

CHAPTER 7

Illocution as a unit of reference for spontaneous speech

An account of insubordinated adverbial clauses in Brazilian Portuguese

Giulia Bossaglia, Heliana Mello and Tommaso Raso
Federal University of Minas Gerais, FAPEMIG, CNPq

In this paper we propose a synchronic, corpus-based account of insubordination, through the analysis of adverbial clauses in Brazilian Portuguese spontaneous speech at the syntax/prosody interface. The segmentation of the speech flow through prosodic cues is crucial to analyse linguistic and, specifically, syntactic relations in spoken language. Besides, it is through prosody that illocutionary and informational values are conveyed in speech. Our claim is that insubordination can be studied without assuming the existence of a grammaticalization path or main clause ellipsis processes, given that through specific illocutionary prosodic profiles, syntactically dependent clauses are assigned pragmatic autonomy.

Keywords: spontaneous speech, speech segmentation, illocution, syntax/prosody interface, insubordination, adverbial clauses, Brazilian Portuguese

1. Introduction

In this paper, spoken language is understood to be a process, rather than a product. In this sense, we assume no bias regarding a formal system that is put to work in the production of well-formed sentences. As spontaneous speech is produced on the go, we claim that prosody is above any other structural level in conveying linguistic meaning and, as we show, syntax does not behave in this realm as it often does, or is assumed to behave, in written texts. The incremental nature of spoken syntax is paired with a necessary prosodic counterpart, constituted by phenomena such as prosodic boundaries and tones, so that whatever is said can be bound to meaning and understood in the specific domain in which it is produced. This process takes into account the information structure through which a given utterance is rendered, including its resulting illocutionary value, both of which are conveyed

https://doi.org/10.1075/scl.94.07bos
Additional files available at http://doi.org/10.1075/scl.94.0307bos/audio

by prosody (Cresti, 1994, 2000, this volume; Moneglia & Raso, 2014; cf. Mithun, this volume, Part I).

Spoken syntax is yet to be well understood, despite the advances in descriptive and explanatory attempts that have been achieved through spoken corpora studies (Blanche-Benveniste, Bilger, Rouget, & Eynde, 1990, on French; Cresti, 2000, on Italian; Miller & Weinert, 1998, on English, German and Russian; cf. Debaisieux & Martin, this volume; Izre'el, this volume, Part I, among many others). Following this recent tradition of empirical linguistics, in this paper we present data on apparent adverbial clauses in spontaneous spoken Brazilian Portuguese (BP) that behave as utterances, aiming to illustrate how spoken syntax breaks away from commonly assumed views about the sentence as the necessary construct in which encapsulated dependency relations have to be fully fulfilled for linguistic meaning to emerge.

In fact, the same locutive content may be interpreted in different ways, depending on the prosodic patterns through which it is performed, which in turn can convey different informational statuses. The main claim of this paper is that, differently from the traditional definition of insubordination as the result of a grammaticalization path (Evans, 2007), insubordinated constructions can be analyzed synchronically: In speech, the illocutionary informational status of adverbial clauses provides these apparently dependent structures with pragmatic autonomy, via prosody.

In Section 2, we provide a brief review of the literature on insubordination. In Section 3, we propose a rationale for the segmentation of speech based on prosodic cues, necessary in order to explain what we mean by informational status, with special attention to the illocutionary one. Section 4 is dedicated to the illustration of the spoken corpus from which our data come from. Data analysis is detailed in Section 5, in which adverbial clauses in spoken BP are examined taking into account the syntax/information structure interface, mainly the illocutionary use of clauses introduced by lexical operators traditionally considered as subordinators. Final remarks are presented in Section 6. Due to the importance of the prosodic dimension for speech segmentation, informational values, and illocution, we provide the audio files for all the examples, which can also be found at <www.c-oral-brasil.org> → Multimídia.

2. Subordination and insubordination

As part of the complex functional arrangements found in speech, the phenomenon now known as insubordination has great relevance as it breaks apart theoretical expectations related to the primacy usually attributed to a predictable, well-behaved syntax. As we will show, sentential patterns in spontaneous speech may not project dependency relations such as embedding phenomena, since spoken syntax interacts with information structure in such a way that assigns to pragmatics a communicatively higher status than syntax.

Insubordination acquired the interest of linguists especially after the publication of Evans (2007), who established the term "insubordination" to refer to the main clause use of constructions that formally would be considered to be subordinate clauses because they portray a complementizer as their initial element. As proposed by Evans, in the diachrony of insubordination there would have been an elliptical process of the matrix clause. This, in turn, leads to the reanalysis of a construction that does not require syntactic dependency, hence the licensing of the apparent insubordinated clause as main clause. The motivation for the path proposed by Evans relies on pragmatic grounds, which seem to be supported by a wealth of research in several languages, even typologically far apart ones, as will be mentioned shortly. Evans' analysis has mapped insubordination into several functions such as the expression of epistemic meaning, desires, requests and warnings, among others. Evans (2009, pp. 1–2) provides some examples of such use in different languages as illustrated in (1) to (3):

(1) Free-standing conditional clause, introduced by *if*, functioning as request:
If you could just sit here for a while please. (English)

(2) Free-standing chained-form verb functioning as informal imperative:
Are wo mi-te! (Japanese)
That ACC look-CNJ
'Look at that!'

(3) Finite subordinate clause, each word of which bears a *complementizing oblique* case suffix marking the clause as the complement of some main predicate:
Kajakaja-ntha dali-jurrk? (Kayardild)[1]
daddy-COBL come-IMM:COBL
'(Have you seen / do you know) whether/that daddy has arrived?'

Evans (2009) points out that

> in principle, any structural feature associated with subordinate clauses may turn up in insubordination, e.g. subordinating conjunctions…, subordinating verbal morphology…, case use characteristic of subordinate clauses…, subordinate-specific word order. (p. 2)

The term *insubordination* had also been used previously by Aviles, Hale, and Salamanca (1987) in a paper about complements in Miskitu (Misumalpan family,

[1] Tangkic family, Australia.

Nicaragua). The authors note that in causative constructions in Miskitu, it is the verb in the formally dependent clause, the so-called effect verb, that has free tense, therefore determining the causative verb tense – the tense assigned to the causative verb is not licensed in root clauses in Miskitu. In this view of insubordination, there is no free-standing subordinate clause, but an inversion in dependency roles, as it is the verb in the dependent clause that seems to have control over that of the main clause verb, as shown in (4) (Aviles et al., 1987, p. 1).

(4) *Yang mita tuktan ba yab-rika kauhw-bia*
I AG child DEF cause-FC fall-FUT
'I will make the child fall'

In (4), the causative verb *yab-aya* 'to cause' is marked with the *future connective* form, which is a dependent verb form according to the authors. Hence, in this view, as well as in Evans', an unexpected role is performed by what would, otherwise, be considered a dependent clause.

Following Evans (2007), several publications exploring insubordination in different languages, discussing its emergence through language change and grammaticalization processes have appeared, among which there are, among others, Debaisieux (2013), on French *parce que, puisque, quand, si, alors que, tandis que, que* clauses; Gras (2011, 2013), on Spanish free *that*-clauses; Inbar (2016), on *ʃɛ* 'that' clauses in spoken Hebrew; Mithun (2008), on Amerindian languages such as Navajo; Sansiñena, De Smet, and Cornillie (2015), on the developmental path from subordination to insubordination in English, French, German, and Spanish data; Van Linden and Van de Velde (2014), on the diachrony of autonomous and semi-autonomous subordination in Dutch; and Wide (2014), on Swedish *att* 'that' clauses.

The occurrence of free "dependent clauses" in speech, however, had been noticed and studied by several scholars prior to the emergence of the term "insubordination" to refer to such phenomenon as, for example, Lombardi Vallauri (2004, 2010, 2016), on free conditionals in different languages; Mithun (2005), on independent uses of dependent structures in Hualapai (Hokan family, Arizona); Schwenter (1996, 2016a, 2016b), on free conditionals in Spanish. The most comprehensive and up-to-date collection of studies on insubordination to this day is Evans and Watanabe (2016), which approaches a great variety of theoretical issues related to the topic on languages from all the continents.

Shifting our focus to Brazilian Portuguese, Decat (1993, 1999, 2001, 2004) refers to independent and semi-independent subordinate clauses as "loose sentences" and "independent utterances" in both spoken and written registers. Through a grammaticalization process, these structures are assumed to be licensed by focalization and argumentative strategies. Even though prosody is not taken into account, the author points out that, in speech, such loose sentences occupy "dedicated intonation units" (cf. the notion of "idea units" in Chafe, 1988, 1994). Decat states that the informational value of an insubordinated construction is very high and corresponds to a speech act that functions as a focalizing strategy, much like a cleft construction, to bring to the fore an assertion, or an argument relevant to the ongoing interaction.

As we show in the following sections, our theoretical point of view differs from Decat's in as much as we rely on prosodic parameters to segment speech and clearly state what our units of analysis are, recognizing utterances, defined below, as the reference unit for speech. Therefore, it is only through the analysis of the sound signal that one can decide whether a certain speech stretch comprises one or more information units (IUs) and what their status is *vis-à-vis* the utterance. Therefore, our analysis is based on the examination of a spontaneous speech spoken corpus, which has been prosodically segmented and informationally annotated, considering insubordination at the interface between syntax, information structure, and prosody.

3. The analysis of spontaneous speech

3.1 The segmentation of speech

It is only through the accurate segmentation of speech that it is possible to study the linguistic relations among (sequences of) lexical items, as shown in (5).

(5) *PAU: *Não // tá dando a altura daquele que a <Isa>*
NEG is reaching the height of that one that Isa
marcou <lá>/ né //
marked there isn't it
'No // it's reaching the height of that one that Isa marked there / isn't it //' [source: bpubdl01[15]

Reading the sequence in (5), hence without taking into account the prosodic cues, two different interpretations would be equally possible, depending on the preferred segmentation. Actually, since in BP negation can occur in pre-verbal position, we could interpret the sequence as a negative assertion (*it is not reaching the height Isa marked there*) or as two different utterances, a refusal (*no*) plus a positive assertion (*it is reaching the height Isa marked there*). What determines whether the negation must be interpreted as compositional with the subsequent verb or as a different utterance (therefore establishing a different domain) is prosody, which provides the necessary features for the speech stretch to be segmented. By listening to the acoustic signal, it becomes clear that we have to interpret the sequence as two utterances (audios 5a, 5b). Note that there is no pause at the utterance boundary.

There are, of course, different proposals that account for speech segmentation into reference units above the level of the word (turn, spoken sentence, stretch of speech between two pauses – see Cresti & Gramigni, 2004). We propose that accurate segmentation is governed by complex prosodic criteria and that utterance boundaries in many cases do not coincide with a pause, while pauses (even long ones) can appear within utterances, as it has been shown also by statistics-based studies on spoken corpora (Raso, Mittmann, & Oliveira, 2015). In fact, different prosodic cues partake in marking what is perceived as an intonation unit boundary (see Barth-Weingarten, 2016, pp. 13–58, for a survey). As we will explain below, the utterance boundary is also an intonation unit boundary, since utterances can be built up by one or more intonation units. Phonetic cues that characterize boundaries have not yet been completely understood. It is likely that many cues, such as pause, pre-boundary lengthening, f0 reset, change in speech rate and intensity, among others, play a role in marking the perception of a boundary (Mittmann & Barbosa, 2016). It is also necessary to distinguish between at least two kinds of boundaries, terminal and non-terminal, that is, the utterance boundary, perceived as conclusive, and boundaries between different intonation units within the same utterance, perceived as continuing. Intonation units strongly correlate with information units (Chafe, 1994), and they feature specific prosodic profiles that convey their functions.

3.2 Speech segmentation and illocution

We define the utterance as the minimal speech stretch that has pragmatic and prosodic autonomy (Cresti, 2000), and we identify in the utterance the linguistic counterpart of what, since Austin (1962), is called a speech act. Therefore, the utterance conveys an illocution and is delimited by terminal boundaries. It can be made up by one or by more than one intonation units. The only unit which is necessary and sufficient to build an utterance is the one that carries the illocutionary force (corresponding to a *root* unit in the IPO framework: t'Hart, Collier, & Cohen, 1990); other units are optional and carry other informational values (Cresti, 2000; Moneglia & Raso, 2014). It is the illocutionary unit that conveys pragmatic autonomy, that is, the communicative interpretability, to the utterance. There are no morpho-syntactic constraints for the fulfillment of the illocution since it is prosodically licensed (Hellbernd & Sammler, 2016). It has been argued that a significant percentage of utterances in speech, at least a third, does not feature a verb (Biber, Johansson, Leech, Conrad, & Finegan, 1999, p. 1071, on spoken English; Cresti et al., 2004, on spoken French, Italian, Spanish and European Portuguese); these numbers grow up to more than 50% if we include verbal utterances where the verb does not constitute the nucleus; especially in dialogic spontaneous speech, single word utterances are frequent, and even utterances fulfilled by just one interjection or paralinguistic sounds are possible, if they are performed with the appropriate intonation and convey an illocutionary value (Biber et al., 1999; Cresti, 2005; Raso & Mittmann, 2012). The following example (Example (6); audios 6, 6a, 6b; Figure 1) illustrate both the segmentation criteria and the individualization of the illocution.

(6) Vet student talking about challenges and difficulties in trimming a horse's nails:
*LYN: *I mean / they are still long // when I get done with them //*
[source: afammn01[34–35]

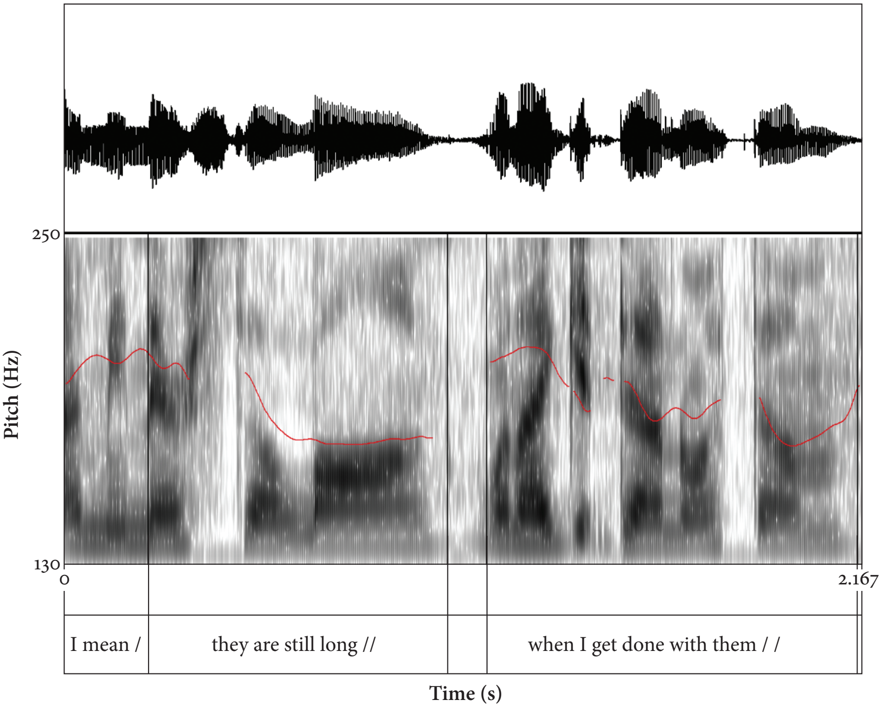


Figure 1. Prosodic contour of Example (6), where two intonational prominences can be seen. The audio files allow their recognition as two different illocutionary forces

As the audio files 6 (a, b) show, the sequence in (6) features two utterances, conveying two similar illocutionary values, pertaining to the representative class and that could be labelled as “conclusion” (Moneglia & Raso, 2014, p. 477). Since the time adverbial clause that corresponds to the second one (audio 6b) is interpretable in isolation due to its illocutionary force, it would be difficult to consider it a subordinate clause that needs, for its interpretation, a main clause as its nucleus. As a matter of fact, the pragmatic autonomy of the utterance is completely independent of the specific illocutionary *values* that a speech stretch can convey, since it is the presence of illocutionary *force* that provides its pragmatic autonomy.

Examples (7) and (8) in BP, with their respective audios (7, 7a, 7b; 8, 8a, 8b), can help the appreciation of the differences between illocutionary and non-illocutionary sequences introduced by the same operator (*que* ‘that’), usually considered as a subordinator:

(7) Man explaining to a friend how voltage of electric current is related to sockets:
*BAL: *cê tá com um jarro d’água // que tem uma espessura assim //*
‘you hold a water jar // that it’s thick like this //’
[source: bfamdl02[61]

(8) *BAL: *tá saindo de uma garrafinha que tem um bico muito pequeno //*
‘it’s coming out from a little bottle that has a very small neck //’
[source: bfamdl02[64]

Examples (7) and (8) apparently feature the same syntactic structure, that is, what would be considered a relative clause introduced by *que*. Nevertheless, in (7) we find two illocutionary sequences, while in (8) just one. This correlates with different segmentations, featuring two utterances in (7) and just one in (8). The different informational status and segmentation of two performances of the same syntactic structure require explanation: They seem to involve the informational/syntactic interface, and prosody as the main linguistic mark. In (7), the clauses can both be interpreted in isolation (audios 7a, 7b), for they convey two illocutions, pertaining to the representative class and that could be labelled as a type of “assertion” the speaker performs to introduce a specific domain of reference in his discourse (he’s using a metaphor between water and electric current). In (8), this is not possible (audio 8a), since there is only one interpretable illocution (independently of its specific value, which could be labelled as another type of “assertion”). Notice that in (8) what seems to be necessary for prosodic and pragmatic interpretation is found within the relative clause (audio 8b) and not the main one. This is so because the *illocutionary* nucleus rests on a few syllables at the right side of the sequence. Actually, the illocutionary information is usually prosodically conveyed by only a few syllables of the intonation unit (for the relation between prosody and illocution, see Cresti, 2018; Moneglia, 2011; Moraes & Rilliard, 2014; Rocha & Raso, 2016).

Example (9) shows an excerpt of 11 utterances, seven of which are verbless.

(9) *KEN: I forget what they call it // the central [/1] little central plaza area //
*LEN: la plaza // mercado / or what //
*KEN: &w [/1] &he / I forget // there was some term they used <for> +
*LEN: <oh yeah> //
*KEN: the [/1] the / <Zocalo> //
*JOA: Zocalo //
*KEN: <the Zocalo> //
*JOA: <the Zocalo> //
*LEN: hum hum // [source: afamcv01[3–14]

Example (9) presents the same locutive content (*the Zocalo*) performed with different illocutionary values (possible labels: "identification" and "conclusion", respectively, for the two utterances of the two speakers), showing that the specific illocutionary meaning does not depend on semantic and morpho-syntactic features but rather on prosodic ones. The speech overlapping makes the analysis of this case difficult; however, Example (10) renders this concept even clearer, as the word *Urano* is performed four times with different illocutionary profiles (respectively: "confirmation", "expression of disbelief", and two instances of "confirmation"), as Figure 2 shows.

(10) *KAT: *o quê //*
'what //'
*SIL: *copos // copos de Urano / que tem aí //*
'glasses // glasses from (*or* made of) Uranus / that are there //'
*KAT: *copos de quê //*
'glasses made of what //'
*SIL: *Urano //*
'Uranus //'
*KAT: *Urano //*
'Uranus //'
*SIL: *é // Urano // Urano //*
'yes // Uranus // Uranus //' [source: bfamdl04[99–107]

In (10), four different utterances with the same locutive content (*Urano*) are uttered by the two speakers. In Figure 2, we can observe how distinct the intonational contours are (probably the main prosodic feature that marks the illocution) for the different utterances. We can also observe that the third and the fourth ones differ

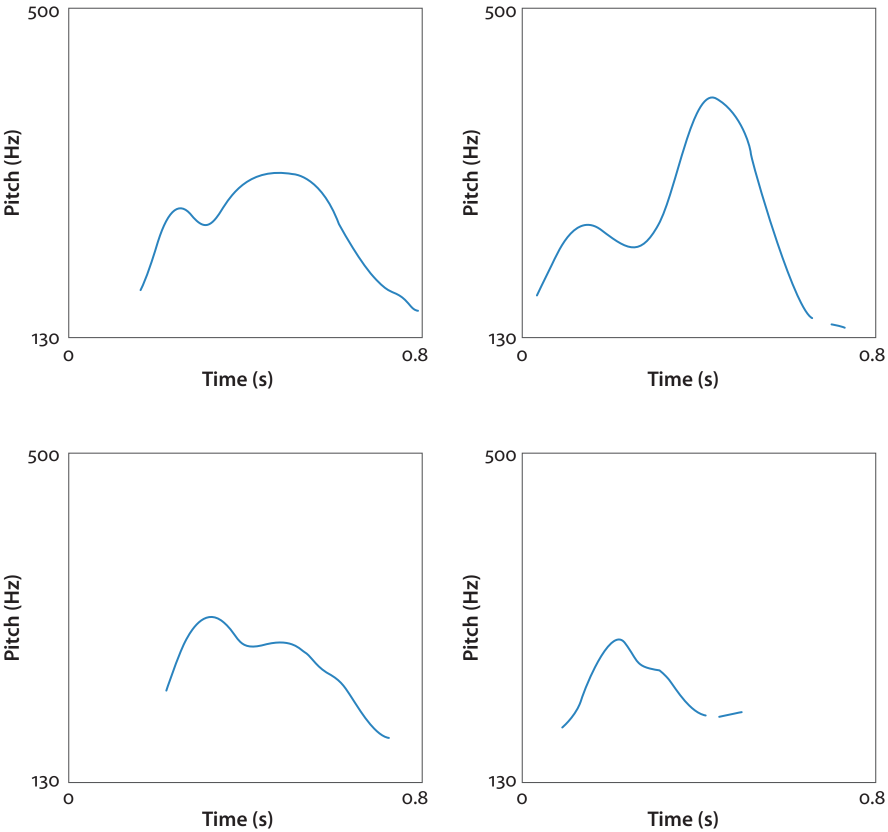


Figure 2. Intonational contours for the different realizations of *Urano* in (10)

mainly in range, but not in form, probably due to a difference in attitude but not in illocution (for the difference between the concepts of illocution and attitude, see Mello & Raso, 2011; Moraes & Rilliard, 2014; Raso & Rocha, 2017).

These segmentation criteria have been adopted for Italian, European Portuguese, Spanish and French in the C-ORAL-ROM corpora (Cresti & Moneglia, 2005), and for BP in the C-ORAL-BRASIL corpus (Raso & Mello, 2012; see also Mello, 2014).[2]

[2] Part of the *Santa Barbara Corpus of American English* (Du Bois, Chafe, Meyer, Thompson, & Martey, 2000–2005) was also segmented and tagged according to the same criteria (Cavalcante & Ramos, 2016).

3.3 Speech segmentation and information units (IUs)

Our data come from a minicorpus extracted from the Informal section of the C-ORAL-BRASIL corpus (Panunzi & Mittmann, 2014), with some examples taken from the minicorpus (Cavalcante & Ramos, 2016) extracted from the Santa Barbara Corpus of Spoken American English (Du Bois et al. 2000–2005). The minicorpora are provided with informational annotation: Each intonation unit was manually tagged for the corresponding IU according to its functional, prosodic and distributional features, following the Language into Act Theory framework (L-AcT; Cresti, 2000; Moneglia & Raso, 2014). Accordingly, two different kinds of IUs can be found: textual IUs and dialogic IUs. Textual IUs partake of the semantic and syntactic text of the utterance, while dialogic IUs are directed to the interlocutor and do not build the semantics of the utterance, corresponding to what, in different frameworks, are called Discourse Markers (Raso, 2014; Raso & Vieira, 2016).[3]

Textual IUs include: (1) Comment (COM), which is the unit that conveys the illocutionary force; (2) Topic (TOP) (Cavalcante, 2015; Firenzuoli & Signorini, 2003; Mittmann, 2012; Raso, Cavalcante, & Mittmann, 2017), defined as the cognitive domain of application of the illocutionary force (i.e., between Topic and Comment there is a relation of pragmatic aboutness; when there is no Topic unit, the illocution is "unloaded" on some given element in the context); (3) Parenthesis (PAR) (Tucci, 2004, 2010), which expresses comments about how to interpret the utterance or part of it; (4) Appendixes, which integrate the text of the Comment (APC) or the Topic (APT), and (5) Locutive Introducer (INT) (Giani, 2004; Maia Rocha & Raso, 2011), which introduces a meta-illocution corresponding, mostly, to reported speech.

Besides these IUs, in the examples Multiple Comments (CMM) and Bound Comments (COB) appear as well. They are two other kinds of illocutionary units. Differently from Comment, Multiple Comments present two (or more) patterned illocutions, that through their pattern build a holistic meaning (comparison, list, reinforcement, etc.); Bound Comments, in turn, exhibit a continuation prosodic signal, marking that there is no terminal break between them, and that other non-patterned illocutionary force(s) will be performed before the terminal break occurs (Cresti, 2009).[4] Each IU features its specific prosodic form, which

[3] Dialogic Units are not important for our point here, but they appear in the examples. They are: Incipit (INP), Conative (CNT), Allocutive (ALL), Phatic (PHA), Discourse Connector (DCT) and Expressive (EXP).

[4] In L-AcT intonation units without informational value are called Scanning (SCA) units; they are part of a bigger IU that is performed through more than one intonation unit, for stylistic or performance related reasons.

is considered the linguistic marker of the function, and its distribution with respect to the Comment, which in turn is the only distributionally free unit. In the case of Comment, the prosodic form varies depending on the illocution; however, Comment always features a prosodic prominence that constitutes its functional nucleus and is aligned with the syllables conveying the prosodic interpretability of the illocutionary force (Moneglia, 2011; Moraes & Rilliard, 2014; Raso & Rocha, 2017; Rocha, 2016; Rocha & Raso, 2016).[5] The Topic form has a functional prosodic nucleus on its right, and its distribution is always at the left of the Comment; the other units have holistic forms (i.e., they do not feature any functional nucleus), each one with peculiar prosodic cues and their own distributions.

In this framework, syntactic relations are seen as subordinated to informational relations. According to Cresti (2014), the scope of true syntactic relations is local, corresponding to the domain of a single IU, considered as a semantic and syntactic island, while no semantic or syntactic compositionality is found between the locutive content of different IUs. The relations among IUs would not have a syntactic nature but a functional one, conveyed by prosody. Thus, it is claimed (Cresti, 2014) that there is no syntactic compositionality in cases like (7) above, or between Topic and Comment, as in examples like (11)–(13). In all cases, the semantic relation between Topic and Comment is understood as conveyed through a prosodic pattern.

(11) *ALI: the scene of the opera /=TOP= New York /=CMM= in eighteen-seventy //=CMM= [source: afamcv05[24]

(12) *ALC: my new boss /=TOP= she came [/2] she told yyy yesterday /=INT= she's /=INT= I wanna be there at seven o'clock to go /=SCA= to community meeting //=COM= [source: afamdl03[1]

(13) *RIC: and /=DCT= the other architect /=TOP= is his nephew or something //=COM= [source:afamdl01[100]

In (11) and (12) it is not possible to reconstruct the syntactic relation between Topic and Comment: In (11) because there is no verb; in (12), because the verb in Comment has already a subject. On the contrary, in (13) we could say that it is possible to reconstruct the syntactic relation between Topic and Comment and that the Topic is the syntactic subject of the verb in Comment.

However, this does not change the fact that both in cases like (11) to (12) and (13) the semantic relation between Topic and Comment is conveyed by a prosodic prominence. In order to show it in a clearer way, we edited the audio file for

[5] For the concept of Focus as the informational functional prominence, see also Cresti (2011).

Example (14), cutting out all the syllables preceding the Topic functional nucleus (see Figure 3, where the nucleus of the Topic is circled). The result is that both audios 14 and the edited 14a allow perceiving the semantic and pragmatic relation between the two IUs. 14a makes it manifest that only a few nuclear syllables of the unit are necessary in order to convey the prosodic pattern, which, in turn, is responsible for marking the semantic and pragmatic relation with the Comment, independently of any syntactic interpretations.

(14) *SHE: *a orientadora* /=TOP= *ela não quer fazer o papel da coordenadora* //=COM=
'the advisor / she doesn't want to assume the role of coordinator //'
[source: bpubmn01[72]

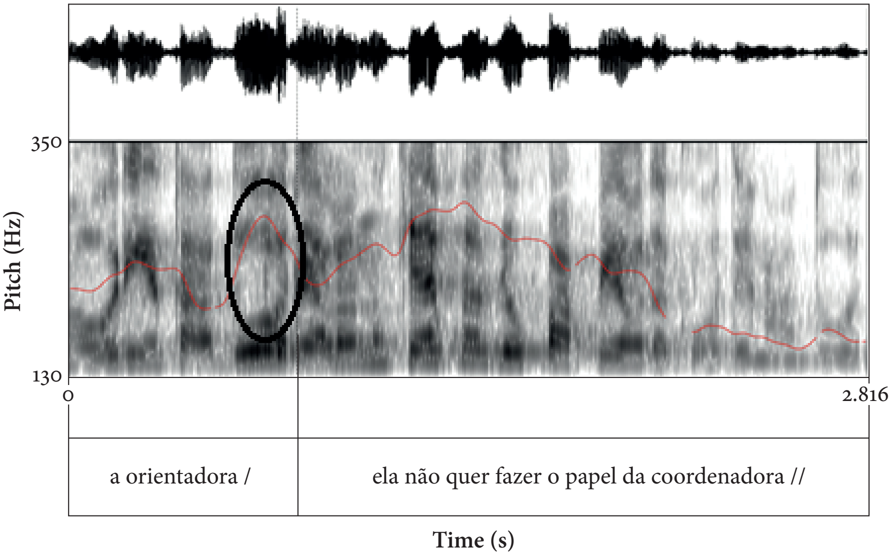


Figure 3. Prosodic contour of Example (14). The Topic functional nucleus is circled

From our perspective, in the Topic-Comment pattern (as between other IUs) the functional relations are firstly conveyed by prosody, therefore accounting for cases like (11)–(12), which present no direct syntactic relations. However, it seems that this does not as yet allow for the conclusion that there is no syntactic processing whenever it is possible to reconstruct some syntactic compositionality between different IUs, as in (13), which still constitutes the majority of the Topic-Comment cases. Evidence (mainly of psycholinguistic nature) is still needed for a firm assertion regarding the reconstruction of syntactic relations across functional prosodic breaks to be made.

We will see other consequences of the informational/syntactic interface in the data discussion.

4. The corpus

The BP minicorpus is a representative sample of the informal section of the C-ORAL-BRASIL corpus (Raso & Mello, 2012). It comprises 20 recording sessions (28,457 words; 5,484 utterances), the transcripts with prosodic boundaries annotation, the audio files, and the text-to-speech alignment (through the WinPitch software: Martin, 2004). The BP minicorpus follows the same architecture of the C-ORAL-BRASIL in what regards the communicative contexts (private/familiar vs. public) and the proportions between monologic (1/3) and dialogic (2/3) interactions (see Panunzi & Gregori, 2011; Panunzi & Mittmann, 2014, for details). The informational tagging was made manually, and it follows the L-AcT framework.

The syntactic phenomenon at issue was explored through the study of finite adverbial subordination, so that the data were collected searching for the occurrences of the adverbial subordinators in the minicorpus, in order to find formally dependent adverbial clauses.

5. Data analysis

5.1 Types of adverbial conjunctions

A rich inventory of canonical and non-canonical adverbial conjunctions was retrieved, as shown in Table 1. The semantic values of the subordinators are displayed according to their traditional descriptions, keeping in mind that defining adverbial clauses on the basis of the presence of adverbial subordinators is quite a circular way to look at them (Kortmann, 1997, pp. 56–57), since the semantic relation between main and subordinate clauses is not determined by the presence of such specific morphemes (Cristofaro, 2005, p. 155). Nonetheless, almost all the single-word or multi-word expressions considered in this study as adverbial subordinators correspond to what Kortmann (1997, p. 72) defines as "*ideal* adverbial subordinator", based on different European languages. However, we will show that not all these operators are used as true adverbial subordinators in spontaneous speech.

Despite the rich inventory of adverbial subordinators, not all of them share the same frequency within the minicorpus (see Table 2). The most frequent adverbial subordinators are *porque* 'because', *se* 'if', and *quando* 'when', while the other

Table 1. Adverbial subordinators in the BP minicorpus

Adverbial subordinator	**Value**
porque 'because'	Cause/Reason
como 'since'	
já que 'since'	
igual 'as'	Manner
como 'as'	
se 'if'	Condition
caso 'in case of'	
quando 'when'	Time
na hora que 'when'	
enquanto 'while'	
depois que 'after that'	
assim que 'as soon as'	
apesar que 'although'	Concessive
se bem que 'although'	

Table 2. Frequency of adverbial subordinators in the BP minicorpus

Subordinator	**Occ.**	**%**
porque 'because'	160	45.5%
se 'if'	84	24.0%
quando 'when'	51	14.5%
como 'as'	14	4.0%
na hora que 'when'	14	4.0%
já que 'since', *apesar que* 'although', *se bem que* 'although', *assim que* 'as soon as', *enquanto* 'while', *depois que* 'after', *igual* 'as'	28	8.0%
Total	**351**	**100.0%**

subordinators appear with much less frequency within the minicorpus.[6] For this reason, we will focus mostly on the three most frequent adverbial subordinators. Taking a closer look at their distribution within the utterance, it becomes clearer that such subordinators do not always fulfill the function of introducing true or canonical adverbial clauses in spontaneous speech.

[6] The same adverbial subordinator appears to be the most frequent in spoken French, too (Debaisieux, 2013, p. 186, on *parce que*).

5.2 Non-canonical characteristics of adverbial clauses in spoken BP

As it was mentioned, it is the prosodic/pragmatic information that provides autonomy to utterances, rather than the presence of well-formed clauses in their locutive content. This has been observed through the fact that, in spontaneous speech, utterances with a verb as syntactic nucleus usually correspond to 50% up to 70%, while approximately a third part is made up by verbless utterances (cf. 3.2).

Accordingly, canonical adverbial clauses were found in the minicorpus together with non-canonical ones, whose characteristics are briefly illustrated in this section. For a full comprehension of the examples, it is recommended to listen to the audio files. First, it is common to find adverbial clauses with non-canonical matrix clauses (in bold in the examples), as in (15) and (16):

(15) Father reporting on his career experiences to his daughter:
*JOR: *e é um caso interessante nesse mercado* /=TOP= *que muito deles me convidavam pra ser sócio deles* //=COM= ***não sócio no papel*** /=CMM= *porque eu era empregado das multinacionais* //=CMM=
'And it is an interesting case in that business / that many of them kept inviting me to be their partner // ***not formally partner*** / because I was working for the multinationals //' [source: bfammn06[60–61]

(16) Haematologist explains how blood is collected and stored in the hospital:
*BRU: *não tem nada que pode ser aproveitado* //=COM= *se* < *tiver* /=SCA= *qualquer doença* > //=COM=
'is there anything that can be useful // if it has / any illnesses //'
*MAR: *quando tá* < *com sorologia positiva* > /=TOP= ***não*** //=COM=
'when it has positive serology / ***no*** //' [source: bpubcv01[361–362]

In (15) the adverbial clauses modify a negated noun phrase, while in (16) a negation. In such cases, the well-formedness of the utterance is guaranteed by the prosodic information, conveying its information structure, so that the syntactic well-formedness is not mandatory. Nonetheless, the semantic value of the adverbial clauses presented above is maintained.

Another non-canonical characteristic for adverbial subordinators in speech is the possibility not to introduce a finite verb, nor a verb at all (for cases like the examples below, in which no main clause appears in the utterance, see Section 5.3.3.1):

(17) Customer at a shoe store talking to retailer:
*JAN: ***porque senão levar uma roxa*** //=COM= *eu não sei que eu hhh faço com ela* //=COM=
'***because otherwise to buy a purple*** [shoe] // I don't know what to do with it //' [source: bpubdl02[67–68]

(18) Two construction workers planning their work:
*ROG: *eu vou &coloc* [/3]=EMP= *eu vou suspender mais um pouquim aqui* /=CMM= *vou pegar a linha* /=CMM= *e vou colocar por cima* //=CMM=
'I'm putting [/3] I'm rising it a little more here / I'm taking the line string / and I'm putting it over it //'
*PAU: *ah* /=EXP= ***porque senão*** //=COM=
'ah / ***because otherwise*** //' [source: bpubdl01[8–9]

In (17) and (18) two illocutions that we can label as "discredit" (of an option: to buy purple shoes, and to build a wall without using the line string, respectively) are fulfilled, once more, independently from the syntactic and semantic well-formedness of the utterance.[7]

5.2.1 *Apparent adverbial subordinators*

A further distinction must be made with regards to adverbial subordinators in spontaneous speech. Besides the previously mentioned non-canonical forms of adverbial clauses (and of their main clauses), some adverbial subordinators, as well as many subordinating and coordinating conjunctions, are used as pragmatic connectors in spoken language (Cresti, 2005; Raso & Mittmann, 2012): They are not linking a main and a subordinate clause, but are rather used in order to start a turn or utterance, or to link different speech acts within the discourse. This seems to be the case for many instances of *porque*:

(19) Construction worker planning work with a colleague:
*PAU: *também uma carreira de pedra chatinha* /=TOP= *tem que pôr* //=COM= ***porque*** /=INP= *isso aí também é o seguinte* //=COM=
'also a line of flat stone / it must be put // ***because*** / here's the deal //'
[source: bpubdl01[69–70]

In (19) the subordinator is performed at the beginning of the utterance, alone within an Incipit dialogic IU (which has the specific function of starting a turn or an utterance). It is very common that, differently from written language, conjunctions are used in speech with pragmatic functions such as opening of a turn, or to link different utterances/speech acts (cf. *because*-clauses used as utterance extensions in Ford, 1993, pp. 135–136; or within "turn-constructional units": Couper-Kuhlen, 1996, p. 392; Hopper & Thompson, 2008). This specific use of some conjunctions

[7] Within our approach, the identification of an illocutionary type depends both on prosodic characteristics and specific pragmatic-cognitive parameters (Moneglia, 2011; Raso & Rocha, 2017; Rocha, 2016).

has been recognized in different spoken languages (Cresti, 2005, on Italian *perché* 'because', *e* 'and', *ma* 'but', *che* 'that'; Debaisieux, 2004, 2013, on French *parce que* 'because'; Groupe Lambda-1, 1975, on French *car, parce que*, and *puisque* 'because'; Raso & Mittmann, 2012, on BP *porque* 'because', *e* 'and', *mas* 'but', *que* 'that'). We will see more in depth how the relation between different speech acts is marked by *porque* in Section 5.3.2.1.

5.3 Distribution of adverbial subordinators/adverbial clauses

The above-mentioned adverbial subordinators and clauses occur in the following positions within the utterance:

a. Inside the same information unit (IU), together with the main clause (cf. Chafe, 1984, p. 438, on *bound* adverbial clauses; Couper-Kuhlen, 1996, on *because*-clauses without declination reset; Cresti, 2005, p. 241, on *linearized* position);
b. At the beginning of an IU, that is, after a non-terminal prosodic break within the utterance (the main clause is performed in a preceding IU: cf. Chafe, 1984, on postposed free clauses; Couper-Kuhlen, 1996, on *because*-clauses after a *partial* pitch reset);
c. At the beginning of the utterance, that is, after a terminal prosodic break; the main clause is performed within another subsequent IU of the same utterance (cf. Chafe, 1984, on preposed free clauses);
d. At the beginning of the utterance, that is, after a terminal prosodic break; the main clause is performed in a different utterance (cf. Chafe, 1984, on free adverbial clauses with period intonation).

In the following sections, all these distributions are illustrated in detail.

5.3.1 *Adverbial clause in the same IU of the main clause*

We borrow Cresti's term "linearized" position (Cresti, 2005, p. 241) for adverbial clauses, when they are performed together with their main clauses within the same IU, as in (20):

(20) Woman talking about her adoptive daughter:
*CAR: *não falo porque acho muito pesado* //=COM=
'I don't talk about it because I think it's really painful //'
[source: bfammn05[58]

In (20) the Cause adverbial clause is performed within the same IU together with its main clause (position (a)), that is, it corresponds to what Chafe (1984) calls *bound* adverbial clauses (cf. the idea of an integrated pitch contour in Couper-Kuhlen, 1996; Ford, 1993). In such cases, there is a consensus among different studies that a true dependency relation exists between main and subordinate clause: Blanche-Benveniste et al. (1990) consider these cases as instances of what is called "*micro-syntaxe*" (i.e., proper syntactic relations vs. "*macro-syntaxe*", discourse oriented relations; cf. Avanzi, 2007; Debaisieux, 2004, 2013; Debaisieux & Deulofeu, 2004); in her study on spoken English *because*-clauses, Couper-Kuhlen (1996) finds that the absence of declination reset between main and adverbial clause is the prosodic cue of a direct causal relation between them (i.e., at the propositional level; "intonational subordination", p. 402); Debaisieux (2004, 2013) on French *parce que* clauses observes that within a same intonation unit the adverbial subordinator is fulfilling its canonical linking function between two clauses ("*introducteur de séquence régie liée*", Debaisieux, 2013, p. 189); within the L-AcT framework, Cresti (2014) assumes that the domain of proper syntactic relationships in spontaneous speech corresponds to a single IU, which is then the unique *locus* for syntactic and semantic compositionality to exist ("linearized syntax", p. 368).[8]

It is worth noting that such true adverbial clauses are very rare in our data, representing roughly 6% of the total amount of occurrences in BP.[9], [10] The most frequent configurations in which adverbial clauses appear to be performed in spoken BP include positions (b), (c) and (d).

5.3.2 *Adverbial clause in a dedicated IU*

5.3.2.1 *Topic/Comment pattern, Multiple Comments, and Bound Comments*

Different adverbial subordinators display strong preference for positions (b) or (c): Time (90%) and Condition (87%) clauses appear mostly at the beginning of the utterance, in the Topic unit (main clause in Comment), that is, position (c). On the other hand, *because*-clauses in the Topic-Comment pattern are extremely rare (only two occurrences), while they are performed mostly in combinations

[8] All these authors show that traditional syntactic dependency tests prove that there is a true dependency relationship between main and adverbial clauses in this specific configuration.

[9] Interestingly, the opposite trend was observed for complement clauses, in a way that a degree of iconicity between the semantic and syntactic integration of complement vs. adverbial clauses is detectable in the way these subordinate clauses are performed in speech (Bossaglia, 2014, 2015; cf. Foley & Van Valin, 1984, p. 264; Givón, 1991, 2001, p. 40; Haiman, 1983).

[10] The low frequency of this configuration of adverbial clauses in spoken language had already been pointed out by Chafe (1984, p. 444), and it is confirmed by the data on spoken French and Italian in Debaisieux (2004, 2013) and Debaisieux & Deulofeu (2004).

of illocutionary units (Bound or Multiple Comments), that is, position (b), as Examples (21) and (22) show.[11, 12]

(21) Girl reporting on her study-abroad experience to a friend:
*BEL: ***quando eu cheguei aqui*** /=TOP= *todas as minhas calças tinham ficado lá hhh* //=COM
‘***when I arrived here*** / all my trousers had remained there hhh //’
[source: bfamdl02[243]]

(22) Old woman talking about her past to her grandson:
*DFL: *e eu ficava até com uma certa inveja* /=COB= ***porque papai era muito sisudo*** //=COM=
‘and I used to get kind of envious / ***because dad was very serious*** //’
[source: bfammn02[176]]

Several studies on different types of adverbial clauses in spoken and written language have shown that, when the adverbial clauses are performed in a different intonation/information unit from that of their main clauses (see positions (b), (c) and (d) in 5.3), they cease to be canonical subordinate clauses, but rather assume new functions at different levels: discourse-oriented (Blanche-Benveniste et al., 1990; Groupe Lambda-1, 1975; Thompson & Couper-Kuhlen, 2005), pragmatic/speech act oriented (Dancygier & Sweetser, 2005; Ford, 1993; Moeschler, 1996; Sweetser, 1990), or interactional functions (Couper-Kuhlen, 1996; Ford, 1993; Hopper & Thompson, 2008; Thompson & Couper-Kuhlen, 2005).

Positions (b) and (c) correspond to what Cresti (2014, p. 368) calls “patterned constructions”, in which is said that no compositionality exists between the two IUs, hence, between the two clauses: They rather are pragmatically organized according to the different communicative functions conveyed by the information pattern of the utterance, across different IUs. Since this pragmatic relationship is conveyed in the first place by prosody, there is in principle no necessity for dedicated lexical indexes (i.e., subordinating morphemes, in this specific case) to appear in order to codify the adverbial relation between the clauses, as we can observe in cases like Example (23), which are quite common:

[11] In the Topic/Comment pattern, the *because*-clause is performed in the Comment unit, differently from Time and Conditional clauses.

[12] *Since*-clauses, on the other hand, prefer the Topic position. This fact is consistent with the most frequent/unmarked positions of the two different Cause clauses (Dancygier & Sweetser, 2005; Diessel, 2001, 2005; Ford, 1993).

(23) Construction worker talking to a colleague while working:
*PAU: *cê fica abanando a mão toda hora* /=TOP= *eles nũ* [/1]=SCA= *nũ alimentam* //=COM=
'you keep fanning your hand all the time / they [i.e., mosquitos] don't [/1] don't feed //' [source: bpubdl01[97]

It is exclusively through intonation that the semantic relationship (conditional) between the two clauses in (23) is codified, which proves that it is not the adverbial subordinator that conveys such information in speech.

Therefore, it becomes clearer that both the pragmatic level (related to the utterance understood as a speech act, and to its information structure) and the semantic level, conveyed by intonation, are hierarchically superordinate to the syntactic one. Nonetheless, it is interesting to notice that Time and Condition adverbial clauses display a preference to appear in Topic, which has the function of defining the circumstances corresponding to the domain of application of the illocution in Comment: This means that there is consistency between their semantic value (defining temporal circumstances and conditions for the event codified by the main clause) and the pragmatic function of the IU they seem to "prefer".[13]

As for Cause *porque*-clauses, their preference for the postposed position is consistent with their unmarked position with respect to the main clause. In such cases, it is possible to observe the previously mentioned lack of syntactic and semantic compositionality, as in (24)–(26) below. Notice that the main cue to discern whether compositionality exists or not, is always prosodic in the first place.

(24) Woman reporting on her childbirth experience in a car:
*REG: *no carro* /=TOP= *eu ficava* /=INT= *co Haroldo* /=PAR= *corre* /=COM_r= *Haroldo* //=ALL_r= *ô meu Deus do céu* //=COM_r= *pega meu filho na sua mão* //=COM_r= *pega meu filho na sua mão e segura porque* /=INT_r= *Nossa Senhora* //=COM_r= *a siora que é mãe* /=COB_r= *siora sabe* /=COB_r= *sio' pega meu filho* //=COM_r= *pega meu filho e cuida* /=COB_r= ***porque nũ tinha* /=SCA= *outro recurso* //=COM=**
'in the car / I kept [saying] / to Haroldo / run / Haroldo // oh my God // take my son by your hand // take my son by the hand and hold him because / Saint Mary // You are the mother / You know [how to do it] / You take my son // take my son and take care of him / ***because there wasn't* / *any other means*** //' [source: bfammn04[5–11]

[13] cf. the idea that *when*- and *if*-clauses can be used in speech in order to set up background mental spaces (Dancygier & Sweetser, 2005, p. 11), and Haiman (1978) for the overlapping of conditionals and topics across different languages.

The *porque*-clause in the last utterance of (24) is non-compositional with the clause within the other IU, from both a prosodic and a syntactic point of view (see also the agrammatical *consecutio temporum*). In fact, as it is signaled through the tag "_r" ("reported"), this excerpt contains quite a long reported speech, which ends with the reported Bound Comment (COB_r) of the last utterance. There is, thus, a change of illocutionary plan from the meta-illocutionary one of the reported speech to the speaker's.

(25) Old woman talking about her past to her grandson:
*DFL: *que o meu avô* /=TOP= *era de uma família abastada* /=COB= ***porque o professor ia em casa*** /=COB= *nũ ia po grupo não* //=COM=
'that my grandpa / was of a rich family / ***because the teacher went to [his] place*** / he didn't go to the common school //'
[source: bfammn02[53]

In (25), the causal relation between the alleged main and adverbial clauses is not maintained at the propositional level, as it would be the case if the two IUs were semantically compositional. It is, rather, an instance of indirect causality (Couper-Kuhlen, 1996, pp. 403–404), or of what Sweetser (1990, p. 77) calls "causality in the epistemic domain": The speaker acknowledges that what she said in the first IU was *inferred* by her from what is expressed by means of the *because*-clause ([*he*] *was of a rich family* [and I think so] *because the teacher went to his place*). There is no direct causality between the event described by the adverbial clause and the event/state in the main one, but rather an inferential relationship that the speaker explains to her interlocutor, in a way that the causal relation is shifted from the propositional to the epistemic domain.

Nonetheless, plenty of examples were retrieved that do not display such a straightforward lack of semantic and syntactic compositionality:

(26) Old woman talking about her past to her grandson:
*DFL: *eu &f* [/1]=SCA= *tinha uma certa inveja <da Maria Julieta>* /=CMM= ***porque tinha um pai brincalhão*** //=CMM=
'I &f[/1] was a little envious of Maria Julieta / ***because [she] had a funny dad*** //' [source: bfammn02[183]

In (26) it seems possible to recognize that a direct (i.e., at the propositional level) causal relation exists between main and adverbial clauses, although they are performed each one in a dedicated IU. Additionally, in cases such as (27) both direct and indirect cause interpretations are possible:

(27) Old man telling a story about a legendary-like snake that, according to him, used to live in the Minas Gerais State:

*MAI: *no norte de Mina* /=TOP= *tinha esse* [/2]=SCA= *antigamente* /=PAR= *tinha esse tipo de cobra todo* /=COM= *né* //=PHA= *talvez agora já acabou* /=COB= ***porque já desmataram muito*** /=COM= *né* //=PHA=
'in the North of Minas [Gerais State] / there was that [/2] once / there was all that type of snake / you know // maybe now it has already disappeared / ***because they've deforested a lot*** //'

[source: bfammn01[88–89]

It is not completely clear if it is always possible to discard semantic compositionality at the propositional level between main and *because*-clauses when they are performed in separate IUs. In (27) the indirect cause relation (i.e., in the epistemic domain) between main and adverbial clause could be reconstructed, for example, from the epistemic modality index *talvez* 'maybe' within the main clause. Aspects concerning modality will not be deepened in this study, but represent a necessary step to take in the analysis of spoken syntax, as many studies from different perspectives point out (among others, see Avanzi, 2007; Debaisieux, 2013; see Blanche-Benveniste et al., 1990, on the different scopes of modality in the micro- and macro-syntax domains; from a cognitive perspective, see Dancygier & Sweetser, 2005; Sweetser, 1990; for our definition of modality, understood as the *Modum on Dictum*, see Mello & Raso, 2011). Besides this most frequent position (b) for *because*-clauses, which is also considered its unmarked one (cf. Dancygier & Sweetser, 2005, pp. 180–182; Diessel, 2001, p. 445, 2005, p. 454; Ford, 1993, pp. 89–90), this specific subordinator appears consistently at the beginning of the utterance as well, in both languages, as we will show in 5.3.3.

5.3.3 *Adverbial clauses in a dedicated utterance: Insubordination*

Clauses introduced by adverbial subordinators in a dedicated utterance correspond to what we consider insubordinated adverbial clauses, but a couple of distinctions are needed: first, with respect to the typology of the utterance. In fact, it is possible to find adverbial clauses both in simple or compound utterances, that is, in utterances formed by only one IU, the Comment, or by the Comment plus one or more IUs, as shown in (28) and (29) (repeated with more surrounding context in (30) and (31), respectively):[14]

[14] Here, compound utterances formed by Comment plus *dialogic* units will be considered as simple utterances, since dialogic units are never compositional with the semantic and syntactic content of the utterance.

(28) Woman reporting on her childbirth experience in a car:
*REG: ***porque ninguém acreditava*** /=COM= *né* //=PHA=
'***because no one believed it*** / you know //'
[source: bfammn04[62]; audio: Bossaglia_Mello_Raso - Audio 33]

(29) Boy talking to a friend:
*BAL: ***porque eu nunca confundo letras com <informática>*** /=COB= *nũ tem nem como* //=COM=
'***because I never mistake arts with computer sciences*** / it's really impossible //'
[source: bfamdl02[81]

In (28) a *because*-clause is performed in a simple utterance, while in (29) there is an utterance formed by a combination of illocutionary units (Bound Comments). In the latter, more syntactic and semantic material is present within the utterance together with the *because*-clause and the protasis, but none of it could be considered as main clause-like material for the adverbial clause, since its prosodic profile clarifies that it is an asyndetic juxtaposition of a second clause. Therefore, the configurations exemplified above share the fact that a formally adverbial clause is performed without its main clause within the same utterance. There being more semantic/syntactic material or not, in both cases these syntactic structures are given pragmatic autonomy by their illocutionary prosodic patterns. What we consider as insubordination is the pragmatically independent status of formally dependent structures. Such a pragmatic independence is conveyed by prosody in the first place, as explained in the next section.

5.3.3.1 *"Semi"-insubordinated versus fully insubordinated adverbial clauses*

A second distinction could be introduced: Within adverbial clauses that are performed in a dedicated utterance, we can distinguish between cases in which it would be possible to retrieve their "main clauses" in the adjacent linguistic context ("semi"-insubordinated clauses), from the ones in which this is not possible (fully insubordinated).

Let us look again at the previous examples, with a broader context:

(30) Woman reporting on her childbirth within a car:
*REG: *assim* /=INT= *João nasceu dentro do carro* //=COM= *quê* //=COM_r= *menino* /=CNT= ***isso aí foi um acontecimento*** //=COM= ***porque ninguém acreditava*** /=COM= *né* //
'so / João was born inside the car // what // boy / ***that was an event*** // ***because nobody believed it*** / you know //'
[source: bfammn04 [59–62]

In the utterance immediately preceding the "semi"-insubordinated clause (*because nobody believed it*) we can identify some main-clause material (*that was an event*). The causal relation between these two clauses, though, is not at the propositional level (as it happens when main and adverbial clauses are performed within the same IU): Through the *because*-clause the speaker gives an account of why she performed the previous illocution (*that was an event* [and I'm saying it/I can say it] *because nobody believed it*). The same holds for (31):

(31) Boy chatting with a friend:
*BAL: *<o problema> é* /=INT= *eu vou ter que estudar e me atualizar em duas coisas ao mesmo <tempo>* //=COM=
'the thing is / I'll have to study and get updated in two things at the same time//'
*BEL: *<ah>* /=CMM= *tá* //=CMM= *<ah* /=CMM= *mas é yyyy>* +
'ah / ok // ah / but it's +'
*BAL: ***não que isso fosse me confundir*** //=COM= ***porque eu nunca confundo Letras com <informática>*** /=COB= *nũ tem nem como* //=COM=
'***not that this would confuse me*** // ***because I never mistake arts for computer science*** / it's really impossible //' [source: bfamdl02[77–81]

After having expressed his concern about the eventuality of having to study two different things at the same time, BAL rectifies explaining that he wouldn't get confused anyway. The *because*-clause is, in this context, a means to justify the previous speech act (*this wouldn't confuse me anyway* [and I'm saying it] *because I never mistake arts for computer science*).

This kind of shift to the speech act level (Couper-Kuhlen, 1996; Dancygier & Sweetser, 2005; Sweetser, 1990; cf. the notion of *justification énonciative* in Moeschler, 1996, p. 286) is found for the concessive relation (*apesar que* and *se bem que* 'although') in a few cases within the BP data, as in (32):

(32) Customer at a shoe store talking to retailer:
*JAN: *<essa>* ***aqui não fecha no meu pé*** //=COM= ***apesar que meu pé tá meio sujo*** /=COM= *né* //=PHA= *então não fecha* //=COM=
'***this [shoe] doesn't fit my foot*** // ***although my foot is kinda dirty*** / you know // so it doesn't fit //' [source: bpubdl02 [161–163]

It is interesting to note that in (32) the semantic relation between the two events in the utterances in bold can be reconstructed as a direct cause (it is because the foot is dirty that it is difficult to wear the shoe), but a concessive subordinator is used. This is because a concessive relation exists at the speech act level, that is, not between the semantic content of the two clauses, but between the content of the second one and the fact that the first one shouldn't have been uttered: The speaker says that the shoe doesn't fit her foot, and that she is saying it *although* her foot is dirty, so that it is obvious that it does not fit (cf. Couper-Kuhlen & Thompson, 2000; Günthner, 2000, on *obwohl* concessive clauses in spoken German as means to "correct the validity" of a previous speech act).[15]

Protases can be adjacent to utterances that could be recognized as apodoses, as in Example (33):

(33) Retailer of a shoe store talking to a customer:
*EUG: ***se cê quiser comprar as duas*** //=COM= ***eu fico mais feliz*** /=COM= *viu* //=PHA=
'***if you want to buy both*** // ***I'll be happier*** / you know //'
[source: bpubdl02[223–224]

In (33) the protasis is performed with an illocution that we could label as "suggestion" (audio 33a). Its illocutionary force is prosodically conveyed and independent from the presence of the apodosis in the next utterance, in which an assertive speech act is fulfilled. It is worth mentioning that the same protasis, performed with the same intonation (hence, the same illocution), could be found and "work" as well in a different textual and illocutionary context. Nonetheless, from a discourse standpoint, the protasis carries a strong textual relationship with the subsequent apodosis.

We label these adverbial clauses performed in a dedicated utterance, but with a retrievable "main clause" in another, "semi"-insubordinated: From a prosodic and pragmatic perspective, they are completely autonomous and interpretable, but they can display a strong link with the speech act/utterance in which some main clause-like material is retrievable, or display a sort of textual coherence with it within the discourse.

Within our data, the fully insubordinated uses of adverbial clauses seem to be restricted to protases only, as it is shown in (34) and (35):

(34) Man and woman in a car, searching for a street; at one moment, they get to a steep slope:
*ANE: *então é paralela a essa* //=COM=
'so it's the parallel [street] to that one //'
*CES: *<é>* //=COM=
'yeah //'
*ANE: *<então vamo> <subir* /=CMM= *e> olhar quais são* //=CMM=
'so let's go up / and see which are the streets//'

[15] See also Verhagen (2000), who refers these uses as "epistemic concessivity", in the light of a mental space approach.

*CES: *<então>* +
‘so +’
*ANE: *qual é* /=SCA= *a paralela* //=COM=
‘[see] which is / the parallel street //’
*CES: *muito obrigado* /=COM= *dona* //=ALL= *brigado* //=COM=
‘thank you / madam // thank you //’
*ANE: *eh* /=PHA= ***se cê nũ tiver um carrinho que*** **[/1]** ***que sobe aqui*** **//=COM=**
‘well / ***if you don’t have a good car that [/] that climbs here*** //’
*CES: *ahn* //=COM= *é* //=COM= *isso não é muito bom* //=COM=
‘uhm // yeah // this is not very good //’ [source: bfamdl05[31–41]

(35) Couple in a car, the low-battery alarm of the recording mic starts beeping annoyingly:
*LAU: *Luzia* /=CMM= *desativa essa bomba* /=CMM= *pelo amor de Deus* //=CMM=
‘Luzia / defuse this bomb / for God’s sake //’
*LUZ: *nũ tem jeito* //=COM= *dá muito trabalho agora desativar* //=COM= *andando* /=TOP= *nũ dá não* //=COM=
‘no way // it’s too difficult now to defuse it // while going / it’s impossible //’
*LAU: ***mas e se ela explodir*** **//=COM=**
‘***but if it explodes*** //’
*LUZ: *explode não* //=COM= *a programação dela é pra frente* //=COM=
‘it won’t // it is programmed [to explode] later //’
[source; bfamdl03[188–194]

The protases in bold in the above examples represent what we label “fully insubordinated clauses”: Besides the fact that they are prosodically and pragmatically independent (audios 33a and 34a), no main-clause material is retrievable in the adjacent linguistic context. In (34) the protasis carries an illocution interpretable as expression of obviousness, and in (35) as partial question.[16, 17]

Our idea of insubordination relies on the pragmatic/prosodic autonomy of a formally dependent structure in speech. In this sense, “semi”- and fully insubordinated adverbial clauses share the fact that they correspond to speech acts. “Semi”-insubordinated structures are the ones that exhibit a relationship with another contiguous speech act in which an *apparent* main clause is performed, because

[16] See Lombardi Vallauri (2004, 2010, 2016) for a list of the illocutions conveyed by free conditionals in Italian and other languages.

[17] Example (35) shows that in speech the presence of a *wh*-element is not mandatory for partial question: What is carrying this specific illocution is the falling prosodic profile applied to the locutive content.

the relation between the clauses in the different utterances is not of syntactic, but rather pragmatic or discursive "dependency". Fully insubordinated clauses (only apparent protases within our data), on the other hand, do not display such a relationship with some main-clause material in the adjacent linguistic context. We think that this is not comparable to what Evans (2007) considers one of the steps of the grammaticalization and/or constructionalization of insubordinated clauses, that is, the ellipsis of the main clause. First, because we think that it is possible to look at these insubordination phenomena synchronically, focusing on the prosodic/pragmatic organization of speech. Accordingly, from our perspective there is no need for reconstructing a main clause for the insubordinated adverbial clause: Its interpretability does not depend on the syntactic completeness of the sentence, because it relies on pragmatic grounds (see also Simone, 2009, who defends the idea that insubordinated structures gain their autonomy by virtue of their "*force pragmatique*" only, i.e., their illocution).

6. Final remarks

In this study we have approached insubordination focusing on adverbial clauses in spoken BP. In our perspective, the study of insubordination and, in general, spoken syntax, must take into account the prosodic organization of speech. The speech continuum is organized in sequences of utterances/speech acts, whose segmentation, information structure and illocutionary values are all conveyed by prosodic cues. We have shown that in spontaneous speech the use of adverbial clauses as true subordinate clauses is very rare, and that according to specific prosodic characteristics they acquire specific discourse-oriented and pragmatic functions.

On these grounds, we propose a synchronic and pragmatics-oriented definition of insubordination: The illocutionary force, conveyed by prosodic means, provides adverbial (hence, formally dependent) clauses with pragmatic autonomy. Therefore, the lexical operator usually used as a syntactic subordinator must be interpreted as a pragmatic connector: The relation it establishes is not a syntactic one (with respect to the content of a previous clause), but a pragmatic one with respect to the (con)text.

We label as semi-insubordinated clauses the ones which display a strong textual or pragmatic link with another utterance containing what we called "main-clause material". It is the case of the widespread use of *because*-clauses as means the speakers use in order to justify some of their previous speech acts, or of the "semi"-insubordinated concessive clauses, used in order to "correct" the validity of a previous speech act. It seems, then, that in "semi"-insubordinated configurations the subordinators are used in order to mark explicitly the semantics of the relationship between two speech acts, although shifted to domains other than the propositional one.

What we label as fully insubordinated clauses correspond to protases only, within our data. In these cases, there isn't any retrievable main-clause material in the adjacent linguistic context, being the apparent dependent structure completely loose, from a textual point of view.

Acknowledgements

The authors would like to acknowledge CNPq and FAPEMIG for the research grants that have made writing this paper possible.

Abbreviations

ACC	accusative	FC	future connective
AG	agent	FUT	future
BP	Brazilian Portuguese	IMM	immediate
CNJ	conjunct/chained form	IU	information unit
COBL	complementizing oblique	L-AcT	Language into Act Theory
DEF	definite		

C-ORAL notation and symbols list

ALL	Allocutive
APC	Appendix of Comment
APT	Appendix of Topic
CMM	Multiple Comment
CNT	Conative
COB	Bound Comment
COM	Comment
DCT	Discourse Connector
EMP	Empty
EXP	Expressive
INP	Incipit
INT	Locutive Introducer
PAR	Parenthesis
PHA	Phatic
SCA	Scanning unit
[tag]_r	reported [information unit]

TMT	Time Taking unit
TOP	Topic
< >	speech overlapping
hhh	paralinguistic sound (e.g., laughter, whistle, click)
[/n]	retracting (n= number of cancelled words)
/	non-terminal prosodic break (intonation unit boundary)
//	terminal prosodic break (utterance boundary)
+	utterance interruption
&	word interruption (e.g., &interrup)
xxx	unintelligible word
yyyy	unintelligible sequence of more than one word
yyy	censored word
&he	time taking filler
a	American English (e.g., **a**famdl01)
b	Brazilian Portuguese (e.g., **b**famdl01)
fam	private/familiar context (e.g., a**fam**dl01)
pub	public context (e.g., a**pub**dl01)
dl	dialogue (e.g., afam**dl**01)
cv	conversation (e.g., afam**cv**01)
mn	monologue (e.g., afam**mn**01)
01	number of recording file (e.g., afammn**01**)
[01]	utterance number (e.g., afammn01[**01**])
*XYZ	speaker name acronym

References

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

Avanzi, M. (2007). Regards croisés sur la notion de macro-syntaxe. *Revue Tranel* (Travaux Neuchâtelois de Linguistique), 47, 39–58.

Aviles, A., Hale, K., & Salamanca, D. (1987). *Insubordinated complements in Miskitu*. Manuscript, MIT. Retrieved from <http://web.mit.edu/linguistics/halepapers/papers/hale098.pdf>

Barth-Weingarten, D. (2016). *Intonation units revisited. Cesura in talk-in-interaction*. Amsterdam: John Benjamins. https://doi.org/10.1075/slsi.29

Biber, D., Johansson, S., Leech, G., & Conrad, S. (1999). *The Longman grammar of spoken and written English*. Harlow: Pearson Education.

Blanche-Benveniste, C., Bilger, M., Rouget, Ch., & Eynde, K. van den. (1990). *Le français parlé. Etudes grammaticales*. Paris: Editions du CNRS.

Bossaglia, G. (2014). Interface entre sintaxe e articulação informacional na fala espontânea: Uma comparação baseada em corpus entre português e italiano. *Caligrama: Revista de Estudos Românicos*, 19(2), 35–60.

Bossaglia, G. (2015). Pragmatic orientation of syntax in spontaneous speech: A corpus-based comparison between Brazilian Portuguese and Italian adverbial clauses. *CHIMERA: Romance Corpora and Linguistic Studies*, 2, 1–34.

Cavalcante, F. (2015). The topic unit in spontaneous American English: A corpus-based study (Unpublished Master's thesis). Federal University of Minas Gerais, Brazil.

Cavalcante, F., & Ramos, A. (2016). The American English spontaneous speech minicorpus. *CHIMERA: Romance Corpora and Linguistic Studies*, 3(2), 99–124.

Chafe, W. L. (1984). How people use adverbial clauses. *Proceedings of the Tenth Meeting of the Berkeley Linguistics Society*, 437–449. https://doi.org/10.3765/bls.v10i0.1936

Chafe, W. L. (1988). Linking intonation units in spoken English. In J. Haiman & S. A. Thompson (Eds.), *Clause combining in grammar and discourse* (pp. 1–27). Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.18.03cha

Chafe, W. L. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago, IL: The University of Chicago Press.

Couper-Kuhlen, E. (1996). Intonation and clause combining in discourse: The case of *because*. *Pragmatics*, 6(3), 389–426. https://doi.org/10.1075/prag.6.3.04cou

Couper-Kuhlen, E., & Thompson, S. A. (2000). Concessive patterns in conversation. In E. Couper-Kuhlen & B. Kortmann (Eds.), *Cause, condition, concession, contrast: Cognitive and discourse perspectives* (pp. 381–410). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110219043.4.381

Cresti, E. (1994). Information and intonational patterning in Italian. In B. Ferguson, H. Gezundhajt, & P. Martin (Eds.), *Accent, intonation et modèles phonologiques* (pp. 99–140). Toronto: Mélodie.

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

Cresti, E. (2005). Notes on lexical strategy, structural strategies and surface clause indexes in the C-ORAL-ROM spoken corpora. In E. Cresti & M. Moneglia (Eds.), *C-ORAL-ROM: Integrated reference corpora for spoken Romance Languages* (pp. 209–256). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.15.08cre

Cresti, E. (2009). La Stanza: un'unità di costruzione testuale del parlato. In A. Ferrari (Ed.), *Sintassi storica e sincronica dell'italiano. Subordinazione, coordinazione, giustapposizione*. Atti del X Congresso della Società Internazionale di Linguistica e Filologia Italiana, vol 2 (pp. 713–732). Firenze: Firenze University Press.

Cresti, E. (2011). The definition of focus in Language into Act Theory (L-AcT). In H. Mello, A. Panunzi, & T. Raso (Eds.), *Pragmatics and prosody, illocution, modality, attitude, information patterning and speech annotation* (pp. 39–82). Firenze: Firenze University Press.

Cresti, E. (2014). Syntactic properties of spontaneous speech in the Language into Act Theory: Data on Italian complements and relative clauses. In T. Raso & H. Mello (Eds.), *Spoken corpora and linguistic studies* (pp. 365–410). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61.13cre

Cresti, E. (2018). The illocution-prosody relationship and the Information Pattern in spontaneous speech according to the Language into Act Theory (L-AcT). In M. Heinz & M. C. Moroni (Eds.), *Prosody: Grammar, information structure, interaction. Linguistik online*, 88(1), 33–62.

Cresti, E. (this volume). The pragmatic analysis of speech and its illocutionary classification according to Language into Act Theory. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Cresti, E., & Gramigni, P. (2004). Per una linguistica corpus based dell'italiano parlato: Le unità di riferimento. In F. Albano Leoni, F. Cutugno, M. Pettorino, & R. Savy (Eds.), *Atti del convegno "Il parlato italiano"*, Napoli, 13-15/02/2003 CD-ROM. Napoli: M. D'Auria.

Cresti, E., & Moneglia, M. (Eds.). (2005). *C-ORAL-ROM: Integrated reference corpora for spoken Romance Languages*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.15

Cresti, E., Nascimento, F. B. do, Moreno-Sandoval, A., Veronis, J., Martin, P., & Choukri, K. (2004). The C-ORAL-ROM CORPUS. A multilingual resource of spontaneous speech for romance languages. In *LREC* 2004, 575–578. Retrieved from <http://elvira.lllf.uam.es/ING/Publicaciones/coralrom-lrec2004.pdf>

Cristofaro, S. (2005). *Subordination*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199282005.001.0001

Dancygier, B., & Sweetser, E. (2005). *Mental spaces in grammar: Conditional constructions*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486760

Debaisieux, J.-M. (2004). Les conjonctions de subordination: Mots de grammaire ou mots du discours? *Le cas de parce que. Revue de Sémantique et Pragmatique*, 15–16, 51–67.

Debaisieux, J.-M. (2013). *Autour de* parce que *et de* puisque. In J.-M. Debaisieux (Ed.), *Analyses linguistiques sur corpus: Subordination et insubordination en français*, (pp. 185–248). Paris: Hermès-Lavoisier.

Debaisieux, J.-M. & Deulofeu, J. (2004). Fonctionnement microsyntaxique de modifieur ar fonctionnement macrosyntaxique en parataxe de constructions introduites par *que* et *parce que* en français parlé, avec extension au cas de *che* e *perché en italien parlé*. In F. Albano Leoni, F. Cutugno, M. Pettorino, & R. Savy (Eds.), *Atti del convegno "Il parlato italiano*", Napoli, 13-15/02/2003. Napoli: M. D'Auria. Retrieved from <https://halshs.archives-ouvertes.fr/halshs-00149155>

Debaisieux, J.-M., & Martin, Ph. (this volume). Syntactic and Prosodic segmentation in spoken French. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Decat, M. B. N. (1993). "Leite com manga, morre!": da hipotaxe adverbial no português em uso (Unpublished doctoral dissertation). Pontifícia Universidade Católica, São Paulo, Brazil).

Decat, M. B. N. (1999). Por uma abordagem da (in)dependência de cláusulas à luz da noção de "unidade informacional". *Scripta* (*Lingüística e Filologia*), 2(4), 23–38.

Decat, M. B. N. (2001). Orações adjetivas explicativas no português brasileiro e no português europeu: Aposição rumo ao 'desgarramento'. *Scripta* (*Lingüística e Filologia*), 5(9), 104–118.

Decat, M. B. N. (2004). Orações relativas apositivas: SNs 'soltos' como estratégia de focalização e argumentação. *Veredas*, 8(1–2), 79–101.

Diessel, H. (2001). The ordering distribution of main and adverbial clauses: A typological study. *Language*, 77(3), 433–455. https://doi.org/10.1353/lan.2001.0152

Diessel, H. (2005). Competing motivations for the ordering of main and adverbial clauses. *Linguistics*, 43(3), 449–470. https://doi.org/10.1515/ling.2005.43.3.449

Du Bois, J., Chafe, W. L., Meyer, C., & Thompson, S. (2000–2005). *Santa Barbara corpus of spoken American English*, Parts 1–4. Philadelphia, PA: Linguistic Data Consortium.

Evans, N. (2007). Insubordination and its uses. In Irina Nikolaeva (Ed.), *Finiteness. Theoretical and empirical foundations* (pp. 366–431). Oxford: Oxford University Press.

Evans, N. (2009). Insubordination and the grammaticalisation of interactive presuppositions. Paper presented at *Methodologies in determining morphosyntactic change conference*, Museum of Ethnography, Osaka, March 2009. Retrieved from <http://www.r.minpaku.ac.jp/ritsuko/english/symposium/pdf/symposium_0903/Evans_handout.pdf>

Evans, N., & Watanabe, H. (2016). *Insubordination*. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.115

Firenzuoli, V., & Signorini, S. (2003). L'unità informativa di topic: Correlati intonativi. In G. Marotta (Ed.), *La coarticolazione* (pp. 177–184). Pisa: ETS.

Foley, W. A., & Van Valin, Jr., R. D. (1984). *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Ford, C. E. (1993). *Grammar in interaction: Adverbial clauses in American English conversation*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511554278

Giani, D. (2004). Una strategia di costruzione del testo parlato: l' Introduttore locutivo. In F. Albano Leoni, F. Cutugno, M. Pettorino, & R. Savy (Eds.), *Atti del convegno "Il parlato italiano"*, Napoli, 13-15/02/2003. Napoli: M. D'Auria.

Givón, T. (1991). Isomorphism in grammatical code: Cognitive and biological considerations. *Studies in Language*, 15(1), 85–114. https://doi.org/10.1075/sl.15.1.04giv

Givón, T. (2001). *Syntax: An introduction*. Vol. 2. Amsterdam: John Benjamins.

Gras, P. (2011). Gramática de construcciones en interacción. Propuesta de un modelo y aplicación al análisis de estructuras independientes con marcas de subordinación en español (Unpublished doctoral dissertation). University of Barcelona, Barcelona, Spain).

Gras, P. (2013). Entre la gramática y el discurso: valores conectivos de que inicial átono en español. In D. Jacob & K. Ploog (Eds.), *Autour de que. El entorno de que* (pp. 81–112). Bern: Peter Lang.

Groupe Lambda-l (1975). Car, parce que, puisque. *Revue Romane*, 10, 248–280.

Günthner, S. (2000). From concessive connector to discourse marker: The use of *obwohl* in everyday German interaction. In E. Couper-Kuhlen, & B. Kortmann (Eds.), *Cause, condition, concession, contrast: cognitive and discourse perspectives* (pp. 439–468). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110219043.4.439

Haiman, J. (1978). Conditionals are topics. *Language*, 54(3), 564–589. https://doi.org/10.1353/lan.1978.0009

Haiman, J. (1983). Iconic and economic motivation. *Language*, 59(4), 781–819. https://doi.org/10.2307/413373

't Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual study on intonation: an experimental approach to speech melody*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511627743

Hellbernd, N., & Sammler, D. (2016). Prosody conveys speaker's intentions: Acoustic cues for speech act perception. *Journal of Memory and Language*, 88, 70–86. https://doi.org/10.1016/j.jml.2016.01.001

Hopper, P. J., & Thompson, S. (2008). Projectability and clause combining in interaction. In R. Laury (Ed.), *Crosslinguistic studies of clause combining: The multifunctionality of conjunctions* (pp. 99–124). Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.80.06hop

Inbar, A. (2016). Is subordination viable? The case of Hebrew *ʃɛ* 'that'. *CHIMERA: Romance Corpora and Linguistic Studies*, 3(2), 287–310.

Izre'el, S. (this volume). The basic unit of spoken language and the interface between prosody, discourse and syntax: A view from spontaneous spoken Hebrew. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Kortmann, B. (1997). *Adverbial subordination. A typology and history of adverbial subordinators based on European languages*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110812428

Lombardi Vallauri, E. (2004). Grammaticalization of syntactic incompleteness: Free conditionals in Italian and other languages. *SKY Journal of Linguistics*, 17, 189–215.

Lombardi Vallauri, E. (2010). Free conditionals in discourse: The forming of a construction. *Lingvisticae Investigationes*, 33(1), 50–85. https://doi.org/10.1075/li.33.1.04lom

Lombardi Vallauri, E. (2016). Insubordinated conditionals in spoken and non-spoken Italian. In N. Evans, & H. Watanabe, (Eds.), *Insubordination* (pp. 145–170). Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.115.06val

Maia Rocha, B., & Raso, T. (2011). A unidade informacional de introdutor locutivo no português do Brasil: Uma primeira descrição baseada em *corpus*. *Domínios de Linguagem*, 5(1), 327–343.

Martin, Ph. (2004). Winpitch corpus, a text to speech alignment tool for multimodal corpora. In *LREC* 2004, 537–540. Retrieved from <http://www.lrec-conf.org/proceedings/lrec2004/pdf/780.pdf>

Mello, H. (2014). Methodological issues for spontaneous speech corpora compilation: The case of C-ORAL-BRASIL. In T. Raso & H. Mello (Eds.), *Spoken corpora and linguistic studies* (pp. 27–68). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61.01mel

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* (pp. 1–18). Firenze: Firenze University Press.

Miller, J. E., & Weinert, R. (1998). *Spontaneous spoken language: Syntax and discourse*. Oxford: Oxford University Press.

Mithun, M. (2005). On the assumption of the sentence as the basic unit of syntactic structure. In Z. Frajzyngier, A. Hodges, & D. S. Rood (Eds.), *Linguistic diversity and language theories* (pp. 169–183). Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.72.09mit

Mithun, M. (2008). The extension of dependency beyond the sentence. *Language*, 84(1), 264–280. https://doi.org/10.1353/lan.2008.0054

Mithun, M. (this volume). Prosody and the organization of information in Central Pomo, a California indigenous language. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Mittmann, M. M. (2012). O C-ORAL-BRASIL e o estudo da fala informal: Um novo olhar sobre o tópico no Português Brasileiro (Unpublished doctoral dissertation). Federal University of Minas Gerais, Brazil).

Mittmann, M. M., & Barbosa, P. (2016). An automatic speech segmentation tool based on multiple acoustic parameters. *CHIMERA: Romance Corpora and Linguistic Studies*, 3(2), 133–147.

Moeschler, J. (1996). *Parce que* et l'enchaînement conversationnel. In C. Muller (Ed.), *Dépendance et intégration syntaxique: Subordination, coordination, connexion* (pp. 285–292). Tübingen: Max Niemeyer.

Moneglia, M. (2011). Spoken corpora and pragmatics. *Revista Brasileira de Linguistica Aplicada*, 11(2), 479–519. https://doi.org/10.1590/S1984-63982011000200009

Moneglia, M., & Raso, T. (2014). Notes on the Language into Act Theory. In T. Raso, & H. Mello (Eds.), *Spoken corpora and linguistic studies* (pp. 468–495). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61.15mon

Moraes, J. A., & Rilliard, A. (2014). Illocution, attitudes and prosody: A multimodal analysis. In T. Raso & H. Mello (Eds.), *Spoken corpora and linguistic studies* (pp. 233–270). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61.09mor

Panunzi, A., & Gregori, L. (2011). DB-IPIC. An XML database for the representation of information structure in spoken language. In H. Mello, A. Panunzi, & T. Raso (Eds.), *Pragmatics and prosody: Illocution, modality, attitude, information patterning and speech annotation* (pp. 133–150). Florence: Firenze University Press.

Panunzi, A., & Mittmann, M. (2014). The IPIC resource and a cross-linguistic analysis of information structure in Italian and Brazilian Portuguese. In T. Raso, & H. R. Mello, *Spoken corpora and linguistic studies* (pp. 129–151). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61.05pan

Raso, T. (2014). Prosodic constraints for discourse markers. In T. Raso, & H. Mello (Eds.), *Spoken corpora and linguistic studies* (pp. 412–467). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61.14ras

Raso, T., Cavalcante, F., & Mittmann, M. M. (2017). Prosodic forms of the topic information unit in a cross-linguistic perspective. A first survey. In A. De Meo, & F. Dovetto (Eds.), *La comunicazione parlata. Proceeding of the international SLI-GSCP conference*, Napoli, 13–15 June 2016 (pp. 445–468). Roma: Aracne.

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

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* (pp. 177–221). Belo Horizonte: UFMG.

Raso, T., Mittmann, M. M., & Oliveira, A. C. (2015). O papel da pausa na segmentação prosódica de *corpora* de fala. *Revista de Estudos da Linguagem*, 23(3), 883–922. https://doi.org/10.17851/2237-2083.23.3.883-922

Raso, T., & Rocha, B. (2017). Illocution and attitude: On the complex interaction between prosody and pragmatic parameters. *Journal of Speech Sciences*, 5(2), 5–27.

Raso, T., & Vieira, M. A. (2016). A description of dialogic units/discourse markers in spontaneous speech corpora based on phonetic parameters. *CHIMERA: Romance Corpora and Linguistic Studies*, 3(2), 221–249.

Rocha, B. (2016). Uma metodologia empírica para a identificação e descrição de ilocuções e a sua aplicação para o estudo da Ordem em PB e italiano (Unpublished doctoral dissertation). Federal University of Minas Gerais, Brazil).

Rocha, B., & Raso, T. (2016). The interaction between illocution and attitude, and its consequences for the empirical study of illocutions. In C. Bardel & A. De Meo (Eds.), *Parler les langues romanes. Proceedings of the International GSCP Conference, Stockholm, 9–12 April 2014* (pp. 69–88). Napoli: Università degli Studi L'Orientale.

Sansiñena, M. S., De Smet, H., & Cornillie, B. (2015). Between subordinate and insubordinated. Paths toward complementizer-initial main clauses. *Journal of Pragmatics*, 77, 3–19. https://doi.org/10.1016/j.pragma.2014.12.004

Schwenter, S. A. (1996). Sobre la sintaxis de una construcción coloquial: Oraciones independientes con si. *Anuari de Filologia*, 21, 87–100.

Schwenter, S. (2016a). Independent *si*-clauses in Spanish. In N. Evans & H. Watanabe (Eds.), *Insubordination* (pp. 89–11). Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.115.04sch

Schwenter, S. (2016b). Meaning and interaction in Spanish independent *si*-clauses. *Language Sciences*, 58, 22–34. https://doi.org/10.1016/j.langsci.2016.04.007

Simone, R. (2009). Espaces instables entre coordination et subordination. In A. Ferrari (Ed.), *Sintassi storica e sincronica dell'italiano. Subordinazione, coordinazione, giustapposizione*, Atti del X Congresso della Società Internazionale di Linguistica e Filologia Italiana, Basilea, 30/06–03/07, (pp. 119–144). Firenze: Cesati.

Sweetser, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620904

Thompson, S., & Couper-Kuhlen, E. (2005). The clause as a locus of grammar and interaction. *Discourse Studies*, 7(4–5), 481–505. https://doi.org/10.1177/1461445605054403

Tucci, I. (2004). L'inciso: Caratteristiche morfosintattiche e intonative in un corpus di riferimento. In F. Albano Leoni, F. Cutugno, M. Pettorino, & R. Savy (Eds.), *Atti del convegno "Il parlato italiano*", Napoli, 13-15/02/2003 CD-ROM. Napoli: M. D'Auria.

Tucci, I. (2010). Obiter dictum. La funzione informativa delle unità parentetiche. In M. Pettorino, A. Giannini, I. Chiari, & F. Dovetto (Eds.), *La comunicazione parlata. Atti del GSCP* (pp. 635–654). Napoli: Università degli Studi L'Orientale.

Van linden, A., & Van de Velde, F. (2014). (Semi-)autonomous subordination in Dutch: Structures and semantic-pragmatic values. *Journal of Pragmatics*, 60, 226–250. https://doi.org/10.1016/j.pragma.2013.08.022

Verhagen, A. (2000). Concession implies causality, though in some other space. In E. Couper-Kuhlen & B. Kortmann (Eds.), *Cause, condition, concession, contrast: Cognitive and discourse perspectives* (pp. 361–380). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110219043.4.361

Wide, C. (2014). Constructions as resources in interactions: Syntactically unintegrated att "that" – clauses in spoken Swedish. In R. Boogaart, T. Colleman, & G. Rutten (Eds.), *Extending the scope of construction grammar* (pp. 353–380). Berlin: Walter De Gruyter.

CHAPTER 8

Narrative discourse segmentation in clinical linguistics

Mira B. Bergelson and Mariya V. Khudyakova
National Research University Higher School of Economics, Moscow, Russia

This chapter deals with segmentation, definition of basic units and annotation of the first corpus of Russian narratives by individuals with brain damage – people with aphasia and right hemisphere damage – and neurologically healthy speakers. We show that parameters such as pause length and intonation contours cannot be used for segmentation of impaired speech. Instead, they use syntactic criteria for the identification of the basic, or – as they are called in this chapter – *elementary discourse units* (EDUs). The Russian CliPS (Clinical Pear Stories) corpus contains multi-layer annotation of audio- and video-recordings, performed on micro- and macro-linguistic level, and can be used as a source for qualitative and quantitative research on various aspects of speech in aphasia and right hemisphere damage.

Keywords: discourse segmentation, corpus annotation, aphasic discourse, aphasia, Pear stories retellings, Russian

1. Introduction

1.1 Narrative discourse and segmentation

The Labovian analysis, worked out and polished in the 40 years since the first seminal publications (Labov, 1972; Labov & Waletzky, 1967) by many researchers (see Johnstone, 2016, for references), focused on the linguistic forms that would manifest components of the narrative schema: orientation, coda, descriptive or narrative passages; on the means of expressing evaluation in the text of the story, on coherence and its instruments, discourse markers, verb forms, reference maintenance and the like (see Bamberg, 2012, as a recent example). This approach centers on information content (informativeness) measurements, information structure (topic, focus, givenness, etc.), discourse structure, coherence

https://doi.org/10.1075/scl.94.08ber
Additional files available at http://doi.org/10.1075/scl.94.08ber/audio

and cohesion, and the genre structure of a given type of narrative. The more recent tradition of narrative research within the interactive sociolinguistics approach (De Fina & Georgakopoulou, 2012, 2015) shifts from the narrative analysis of texts to the analysis of social practices. Here the focus is on interaction in communication, "dynamic relations between participants in communities, texts, and practices" (De Fina & Georgakopoulou, 2015, p. 18).

In clinical linguistics, the discourse by persons with acquired language disorders due to focal brain damage as well as by persons with psychiatric and neurodegenerative diseases has recently become the object of study. Clinical linguists changed their perspective from assessing segregated linguistic skills in various disorders to the idea that communication skills should be assessed as a whole. The concept of "functional communication" (Holland, 1980, 1982) serves to explain how people with aphasia (PWA) achieve their communicative goals in spite of linguistic difficulties they may experience due to their diagnoses (Meuse & Marquardt, 1985). Thus, clinical research follows in the steps of the general linguistics turn for discourse studies looking into the most important genres of everyday life discourse. Analysis of narratives, conversations, procedural and exposition discourse are widely used as assessment and in many cases rehabilitation means (for a review see Linnik, Bastiaanse, & Höhle, 2015).

Regardless of approach, the first and necessary condition of a spoken discourse study is to create consistent and verifiable transcripts as a foundation for the next steps of analysis. In order to investigate narrative structure, it is important to establish clear rules for the segmentation of the "building blocks" of the narrative, that is, discourse units. Various frameworks use different criteria for segmentation of oral discourse into *speech*, or, as they are called in this publication – *basic* – *units*. These criteria make use of syntactic structure, semantics, or prosody, or their combination to come up with "utterances" (Marini, Andreetta, et al., 2011), "segments" (Passonneau & Litman, 1997), "communicative", or "C-units" (Armstrong, Ciccone, Godecke, & Kok, 2011; Miller, Andriacchi, & Nockerts, 2015), "T-units" (Coelho, 2002), "verbalizations" (Glosser & Deser, 1991), "elementary discourse units (EDUs)" (Carlson & Marcu, 2001; Kibrik & Podlesskaya, 2009; Mann & Thompson, 1988; Taboada & Zabala, 2008), "minimal discourse units (MDU's)" (Degand & Simon, 2005), "discourse constituent units (DCUs)" (Polanyi, 1985), and "analysis of speech (AS) units" (Foster, Tonkyn, & Wigglesworth, 2000). Given such variability of segmentation criteria and the resulting units, one needs to carefully choose the options for segmentation when developing a corpus of spoken narratives.

In this study we describe the segmentation scheme that was used in the development and annotation of the Russian Clinical Pear Stories Corpus (Russian CliPS) (Khudyakova et al., 2016) – a corpus of oral narratives by people with brain damage resulting in speech impairment. The segmentation scheme was developed with regard to the specific features of pathological speech as well as research questions of the Russian CliPS project – analysis of narratives on micro- and macrolinguistic level.

1.2 Challenges for segmentation of pathological speech

Choice of the "basic unit" is defined by the nature of the discourse and research purposes. Based on Chafe (1994, 2008, 2014), and the general primacy of the oral production, it is natural, while creating a corpus of spoken discourse, to look for the verifiable elementary units that can be determined on the basis of acoustics: Through a combination of prosodic features, where falling/rising intonation of the tonal accents, changes in pitch, tempo and frequency, and pauses are among the most important (Raso & Mello, 2014). Such elementary discourse units (EDUs; Kibrik & Podlesskaya, 2009) are more or less well defined in typologically different languages, which concurs their universal status.

However, analysis of intonation has certain drawbacks, especially for the analysis of speech in clinical populations. First, it is very time- and labor-consuming and requires use of special software (such as Speech Analyzer, "Speech Analyzer – SIL Language Technology", or Praat, Boersma & Weenink, 2007). In clinical linguistics, analysis of discourse can be part of language assessment, as in standardized tests (Goodglass, Kaplan, & Barresi, 2001; Swinburn, Porter, & Howard, 2004), so the segmentation and annotation system has to be easily understood and consistently applied by a speech therapist in a relatively short time. However, when the main purpose of discourse annotation is research, this should not be a problem. On the other hand, research and practice in clinical linguistics are closely connected, and theoretical findings can improve current assessment criteria and therapy methods. In this perspective, annotation of pathological speech for research purposes should be in line with the current speech assessment practices.

Still, the major challenges for pathological discourse analysis are rooted in the nature of language impairment. When analyzing discourse of people with aphasia (PWA) and right hemisphere damage (RHD), one must take into account all possible deficits of the speakers. Aphasia results from damage to the language-dominant (usually left) hemisphere, and manifests itself in deficits on multiple language levels: phonetics, lexis, grammar. In different types of aphasia, these aspects of language are impaired to various degrees. For example, in non-fluent types of aphasia, the main underlying deficits are in the word articulatory program (as in efferent motor aphasia) or utterance planning (as in dynamic aphasia), which can result in long pauses, false-starts, and self-repairs, as well as agrammatism. On the other hand, speakers with fluent types of aphasia, although making semantic and phonological mistakes, can produce quite fluent narratives (Akhutina, 2015). As for damage to the right hemisphere, it is known to cause problems with prosody comprehension and production, as well as general impairment of narrative planning (Alexander & Hillis, 2008; Heilman, Leon, & Rosenbek, 2004; Seddoh, 2004). So, in order to develop a corpus of speech by PWA and RHD, we should consider all these specifics and choose an option that would be most suitable for research purposes.

The current version of the Russian CliPS corpus includes discourse samples for 67 speakers, embracing patients with various diagnoses, various types of aphasia plus samples by healthy speakers. The quantitative information on the Russian CliPS corpus is shown in Table 1. Working out a concept of an "all-purpose" discourse unit along with the set of criteria for singling these units out, is a separate task underlying all kinds of segmentation, annotation and analysis of the corpus data.

Table 1. Quantitative data on Russian CliPS 1.0 (source: Khudyakova et al., 2016)

Group		**Narrative length (ms)**	**Pauses (%)**	**Narrative length (words)**	**Narrative length (clauses)**	**Narrative length (utterances)**
Acoustic-mnestic aphasia	Mean	231 196	43	281,1	52,6	45
	Range	85 229–473 025	25–55	76–180	18–84	16–69
	SD	106 700	10	122,2	19,7	16,6
Dynamic aphasia	Mean	406 023	60	220,4	39,8	38,8
	Range	138 096–810 867	29–71	135–371	27–59	26–59
	SD	196 132	13	91,1	9,4	9,9
Sensory aphasia	Mean	275 765	40	346,4	66,3	58,9
	Range	148 023–549 223	24–56	170–631	28–110	25–94
	SD	117 912	9	174,7	29,6	25,6
Efferent motor aphasia	Mean	377 137	45	228,8	49,9	43,8
	Range	167 879–1 107 112	26–72	58–436	14–91	14–64
	SD	275 043	14	119,7	24,4	17,8
RHD	Mean	195 922	49	279	63	55,7
	Range	122 845–427 025	39–65	185–477	32–120	29–105
	SD	147 132	11	133,5	39,2	33,8
Healthy speakers	Mean	152 437	33	269,5	53,7	42,2
	Range	47 389–296 805	17–51	88–405	16–80	9–71
	SD	62 524	9	113,7	21,7	18,4

2. Russian CliPS corpus

2.1 Speakers

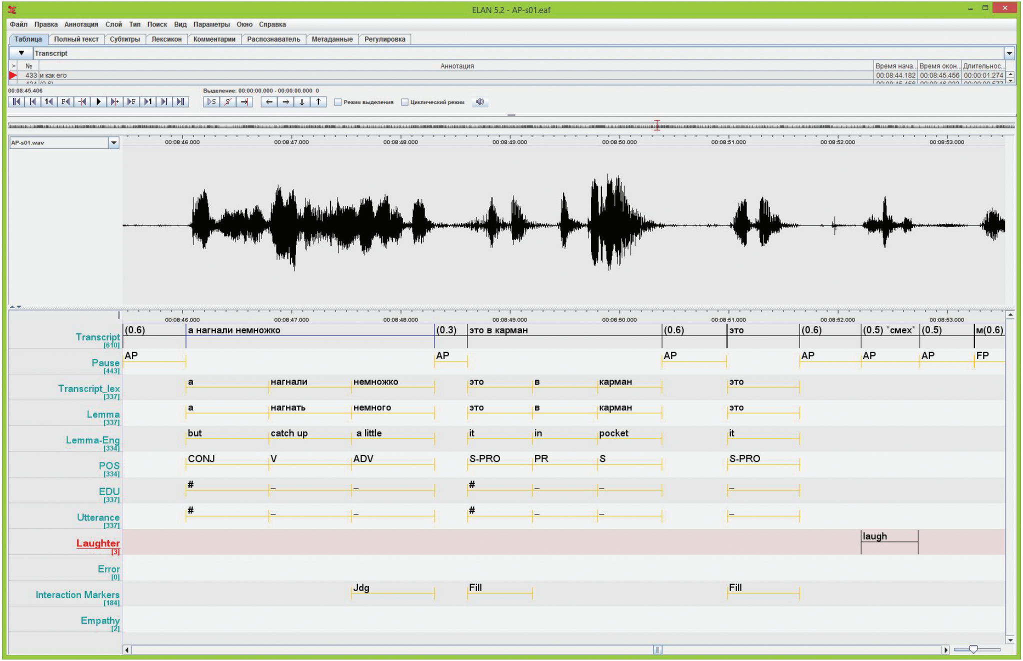
The corpus contains narratives by people diagnosed with chronic aphasia and RHD, and speakers without brain damage. Each speaker from the aphasic group was diagnosed with one aphasia type using Luria's Neuropsychological Investigation (Luria, 1972; see also Akhutina, 2015). The corpus contains 40 stories by people with aphasia of four different types: aphasia with non-fluent speech output (efferent motor aphasia and dynamic aphasia), and aphasia with fluent speech output (acoustic-mnestic and sensory aphasia), (40 people; 17 females; mean age 52.6; range 30–81; SD = 10.5), and five individuals with RHD (5 people; 2 females; mean age 50; range 41–56; SD = 12.3). Speakers from the neurologically healthy group (22 people; 11 females; mean age 58; range 25–84; SD = 13.9) had no history of neurological disease or head traumas. All participants were right-handed, native speakers of Russian language, had at least a high school education, and had normal or corrected-to-normal vision, and no hearing problems.

2.2 Procedure

For the Russian CliPS corpus all speakers were asked to watch the Pear Story film (Chafe, 1980) and then retell it in detail to a person who had not seen it before. The retellings were audio recorded. Video-recordings were made if not objected by the speaker (22 video recordings for the control group and 20 for PWA and RHD).

2.3 Annotation

The annotation of the corpus was performed in ELAN (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006). The annotation scheme includes tiers with transcript, grammatical information, annotation of pause types (filled and absolute pauses) and errors, segmentation into discourse units, and specific tiers for non-verbal sounds and interaction markers (see Figure 1 for a screen caption of the ELAN window with tier annotation).



Transcript

№	Аннотация	Время нача...	Время окон...	Длительнос...
433	и как его	00:08:44.182	00:08:45.456	00:00:01.274

00:08:45.406

Выделение: 00:00:00.000 - 00:00:00.000 0

☐ Режим выделения ☐ Циклический режим

Transcript [610]	(0.6)	а нагнали немножко			(0.3)	это в карман			(0.6)	это	(0.6)	(0.5) *смех*	(0.5)	м(0.6)
Pause [443]	AP				AP				AP		AP	AP	AP	FP
Transcript_lex [337]		а	нагнали	немножко		это	в	карман		это				
Lemma [337]		а	нагнать	немного		это	в	карман		это				
Lemma-Eng [334]		but	catch up	a little		it	in	pocket		it				
POS [334]		CONJ	V	ADV		S-PRO	PR	S		S-PRO				
EDU [337]		#	_	_		#	_	_		_				
Utterance [337]		#	_	_		#	_	_		_				
Laughter [3]												laugh		
Error [0]														
Interaction Markers [184]				Jdg		Fill				Fill				
Empathy [2]														

Figure 1. A sample from the Russian CliPS corpus: ELAN window with tier annotation

3. Segmentation in the Russian CliPS

We define three *types of speech units* in the CLiPS on the basis of the specific linguistic problems and the corresponding analysis that is performed using these units: elementary discourse units (EDUs), utterances, and scenes. EDUs and utterances belong to the so called *basic speech units* (or, *basic discourse units*) which are used in the analysis of narratives at the microlevel, as opposed to scenes that serve as units for the macrolevel analysis. In the following sections, we define these units and provide examples of qualitative analysis performed on each level.

3.1 Basic speech units

3.1.1 *Segmentation criteria*

In aphasia and RHD discourse research, two main approaches to discourse segmentation into basic units co-exist: segmentation based on purely syntactic criteria (Linnik, Bastiaanse, & Khudyakova, 2015; MacWhinney, 2010; Miller et al., 2015), as well as combination of intonation and syntactic criteria (for example, Marini, Andreetta, et al., 2011; Marini, Galetto, et al., 2011; Shewan, 1988). However, both approaches postulate that a unit is supposed to represent a complete thought and correspond to a sentence or a clause in written discourse.

For segmentation into EDUs we have chosen to use grammatical only, rather than prosodic principle. The decision is greatly influenced by the differences in fluency of the speakers, which makes implementation of the acoustic criterion (an easily identified pause and/or pitch or frequency change) problematic.

The amount and duration of pauses varies greatly across speakers of different groups, for example, Figure 2 shows the distribution of absolute (silent) pauses of

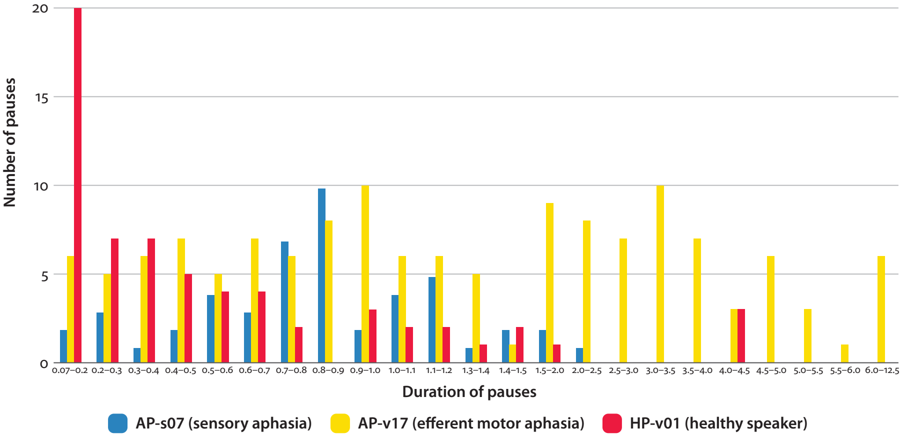


Figure 2. Distribution of absolute (silent) pauses of different durations in narratives by three speakers

different durations in narratives by three speakers: a person with efferent motor aphasia, a person with sensory aphasia and a healthy speaker. While in healthy speech relatively short pauses (less than 1 s) are prevalent with only occasional long ones (1–5 s), in a narrative by a non-fluent speaker pauses with duration ranging from 1 s to 12.5 s are frequent.

In the Russian CliPS corpus, an EDU is defined as a clause with a predicate (see (1a)), or with an omitted predicate (1c), including participle clauses (1b).

(1) a. Healthy speaker (HP-v01)[1]
EDU #52
мимо гордо проходят подростки
mimo gordo proxod-ya podrostk-i
along proudly walk-PRS.3PL teenager-PL.NOM
'Proudly teenagers walk by'

b. Healthy speaker (HP-v01)
EDU #53
жующие его груши
zhuy-ushch-ie ego grush-i
chew-PTCP-PL.NOM his pear-PL.ACC
'Eating his pears'

c. Speaker with efferent motor aphasia (AP-v17)
EDU #39
вот (1.5) *мальчишки ка= ракеткой*
vot (1.5) *mal'chishk-i ka= raketk-oj*
so (1.5) boy-PL.NOM = racket-SG.INST
'So boys <play> with a <tennis> racket'

Each predicate belongs to a separate EDU, except for cases of repetition and word-finding so abundant in clinical discourse. As a result, an EDU with the same information content will greatly range in length in narratives by speakers belonging to the different groups. For example, EDUs in (2a) and (2b) convey roughly the same information content 'the boy takes the basket with pears'. However, in a healthy speaker's narrative (2a) the EDU's duration is 4 s, while in (2b), a narrative by a person with dynamic aphasia (non-fluent type), the EDU lasts significantly longer due to multiple pauses, false-starts and repetitions (39.5 s).

[1] In Examples (1a–1c) the transcript is presented with pauses. The EDU number is provided.

(2) a. Healthy speaker (HP-v05)[2]
EDU #20
ставит (0.1) *корзину* (0.3) *с* *грушами* *себе* (0.1)
stav-it (0.1) *korzin-u* (0.3) *s* *grush-ami* *seb-e* (0.1)
put-PRS.3SG (0.1) basket-SG.ACC (0.3) with pear-PL.INSTR self-DAT (0.1)
на багажник (0.4)
na bagazhnik (0.4)
on carrier-SG.ACC (0.4)
'Puts the basket with pears on his carrier'

b. Speaker with dynamic aphasia (AP-s06)
EDU #11
груши *мальчик* (2.2) *б* (1.4) (2.1) *птс* (0.1) *велосипеде*
grush-I *mal'chik* (2.2) *b* (1.4) (2.1) *pts* (0.1) *velosiped-e*
pear-PL.ACC boy-SG.NOM (2.2) b (1.4) (2.1) pts (0.1) bike-SG.LOC
на велосипеде *мальчик* *маленький* *и* *вот* (3.2) *мальчик*
na velosiped-e *mal'chik* *malen'k-ij* *i* *vot* (3.2) *mal'chik*
on bike-SG.LOC boy-SG.NOM little-SG.M and so (3.2) boy
маленький (3.5) *велосипеде* *и* (1.8) *груши* (5.8) *и* (4.4)
malen'k-ij (3.5) *velosiped-e* *i* (1.8) *grush-i* (5.8) *i* (4.4)
little-SG.M (3.5) bike-SG.LOC and (1.8) pears-PL.ACC (5.8) and (4.4)
мальчик *м* (2.1)
mal'chik *m* (2.1)
boy-SG.NOM m (2.1)
'Pears boy on a bike, on bike little boy, and so little boy on bike pears and boy'

Segmentation of discourse into EDUs based on the syntactic principle lies the foundation for further macrolevel analysis, while still leaving possibilities for the analysis of fluency. For example, analysis of the number of pauses per EDU can serve as a measure of speech fluency, as the data represented in Figures 3a, 3b and 3c demonstrate. For example, we can expect differences in distribution of pauses between fluent (for example, sensory) and non-fluent (for example, efferent motor) types of aphasia.

[2] In Examples (2a–2c) the transcript is presented with pauses. The EDU number is provided.

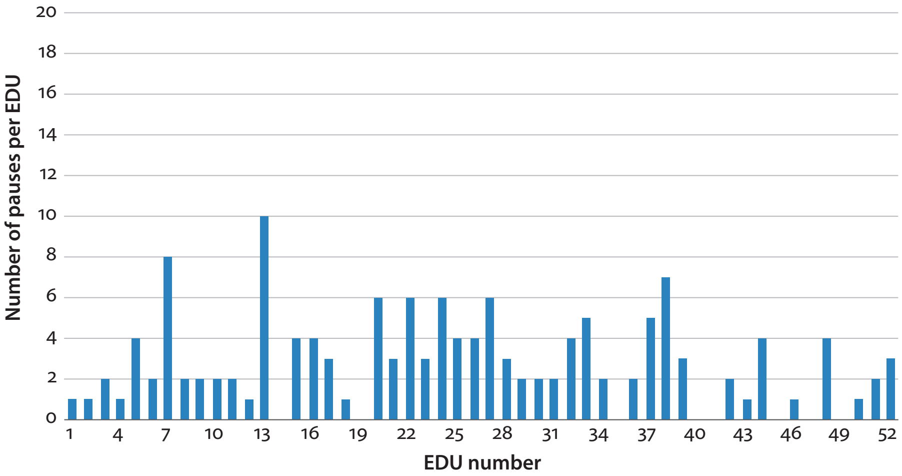


Figure 3a. Number of pauses in EDUs in a narrative by a speaker with efferent motor aphasia (AP-v17)

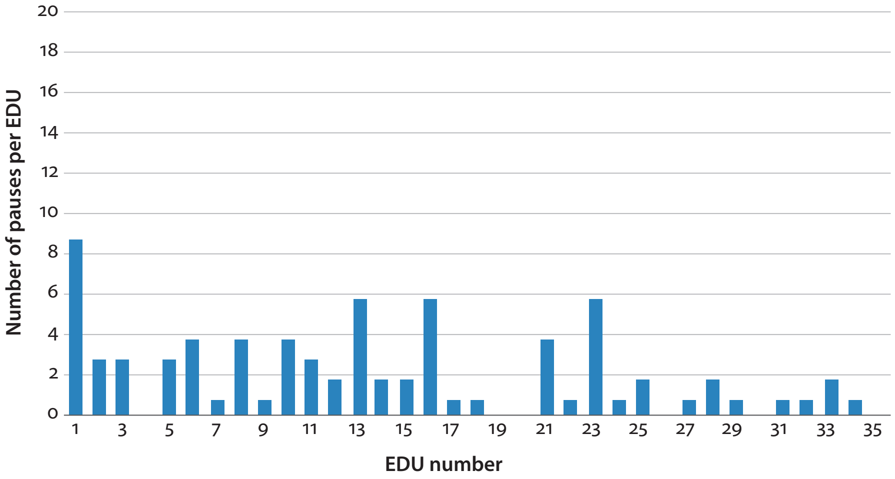


Figure 3b. Number of pauses in EDUs in a narrative by a speaker with sensory aphasia (AP-s07)

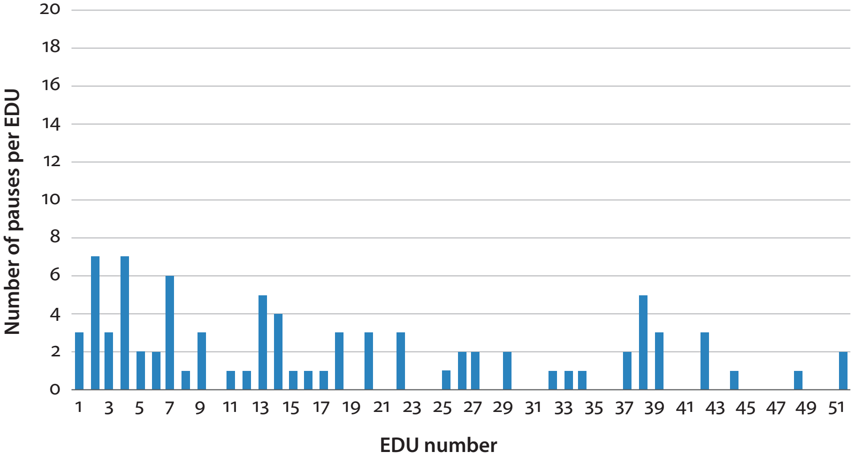


Figure 3c. Number of pauses in EDUs in a narrative by a healthy speaker (HP-v01)

3.1.2 *Basic unit size*

As was discussed above, EDU – the smallest discourse unit in the Russian CliPS corpus – contains one predicate and roughly equals to a clause. However, in clinical linguistics research it is much more common to segment discourse into *utterances*, or as they are also called, C-units (communication units) or verbalizations (Glosser & Deser, 1991; MacWhinney, 2010; Marini, Andreetta, et al., 2011; Miller et al., 2015).

In order to account for grammatical complexity of narratives, we added segmentation into utterances to the corpus annotation scheme. Utterances consist of the main clause with any subordinate clauses. When analyzing discourse, each utterance is given a score based on its syntactic complexity and grammaticality (Glosser & Deser, 1991). However, the notion of an utterance is used only for calculating the amount of clauses per utterance as a measure of syntactic performance of a speaker, and not as a unit for the macrolevel analysis. The other major difference between approaches used in clinical research is the size of the basic unit: whether it roughly equates a clause or a sentence. In this project, our basic units equal clauses.

3.2 Macrolevel segmentation

Segmentation and annotation of the corpus data on the macrolevel is supposed to reveal units to be used for the analysis of the narratives in terms of the adjusted Labovian componential genre schema (Labov, 1997; see also Bergelson, 2007;

Berman, 1997; Polanyi, 1989). Segmentation on the macrolevel focuses on the *scene* as its unit and includes three stages. On the first stage, the transcript of the story is annotated for the *interaction markers* (see (3) below).

3.2.1 *Interaction markers*

The specific situation of storytelling, especially the test-like situation of retelling a video, provokes interaction between its participants: the narrator and the listener. The narrator tells (retells) what happened, describing characters, events and various circumstances, thus creating the *world of the story*. At the same time, the narrator interacts with the listener in the *world of narration* (Norrick, 2000) by addressing the listener, attracting her attention, appealing to her opinion, and also by revealing cognitive and production difficulties.

Linguistic devices that earmark interaction (IMs) may be found inside all types of the discourse passages – descriptive, narrative, instructive, argumentative, and expository. Some of them mark the end or the beginning of the clauses, utterances, or whole episodes. They differ in their specific functions, but all of them serve to signal switching from the world of the story to the world of narration and vice versa.

Introduction of interactional elements in the story schema is much less of the problem for analysis, than for segmentation. The very necessity of postulating non-standard EDUs – truncated, split, and ones without a clear illocution, among others – is related to interactional markers or interactional signals popping up at almost any point in the information flow.

(3) Healthy speaker (HP-v01)[3]

#*мужчина*	*интересного*	*вида*
muzhchin-a	*interesn-ogo*	*vid-a*
man-SG.NOM	interesting-SG.GEN	outlook-SG.GEN

'A man of interesting outlook'

#***да***	*конец*	*лета*	***надо***	***полагать***
da	*konets*	*let-a*	***nado***	***polaga-t'***
well	end-SG.NOM	summer-SG.GEN	must	assume-INF

'**well** (it's) the end of summer **I believe**'

#*собирает*	*урожай*
sobira-et	*urozhaj*
pick-3SG.PRS	harvest-SG.ACC

'picks the harvest'

[3] In Examples (3) and the following, the transcript is presented without pauses. EDU numbers are not provided, a new EDU starts with a # symbol, interaction markers are highlighted with bold, errors are underlined.

#*залез*	*на*	***видимо***	*грушу*	*по*	*приставной*
zalez	*na*	***vidimo***	*grush-u*	*po*	*pristavn-oj*
climb-3SG.PST	on	presumably	pear.tree-SG.ACC	with	standing-SG.DAT

лестнице
lestnits-e
ladder-SG.DAT

'(he) climbed on **presumably** pear tree with a ladder'

The world of narration is conceived as consisting of two separate communication grounds. One is the real situation of storytelling where IMs are directed onto the addressees as part of the conversation. The other is a mental space where the narrators communicate with themselves while planning the next portions of discourse. Doing this they still interact with the listener giving off, revealing their cognitive difficulties with planning, retrieval, or physical production of discourse, see (4):

(4) Speaker with efferent motor aphasia (AP-v09)

#*одна*	***нет***	*ещё*	*пустая*	*почти*	***значит***
odn-a	***net***	*eshche*	*pusta-ja*	*pochti*	***znachit***
one-SG.F.NOM	no	still	empty-SG.F.NOM	almost	so

'One **no** still empty almost **you-know**'

#*во*	*вторую*	***значит***	*в*	*одну*	он	кладёт
vo	*vtor-uju*	***znachit***	*v*	*odnu*	on	klad-et
in	second-SG.F.ACC	so	in	one-SG.F.ACC	he	put-3SG.PRS

та	*груша*	*груша*
ta	*grush-a*	*grush-a*
that-SG.F.NOM	pear-SG.NOM	pear-SG.NOM

'in the other **you-know** in it he puts that pear, pear'

#*которую*	*упала*
kotor-uju	*upa-l-a*
which-SG.F.ACC	fall-PST-SG.F

'which fell'

The interaction components are annotated as a separate annotation layer (see Figure 1). Everything that pertains to the world of storytelling where the speaker interacts with the listener is regarded as interaction, namely: fillers, word search, false starts, feedback markers, appellations to the listener, repetitions, and other discourse markers. Interaction components can occur both within the clause (e.g., *let's call the boy Vovochka*, where *let's* is an element of interaction within a descriptive clause) or comprise a separate clause (e.g., *if you say so*).

At the same time, interaction markers as a class, or more accurately – a set of markers of the storytelling situation dynamics – do not constitute separate discourse units. At least not at the macrolevel of the narrative analysis. And even there, only so called clausal interaction markers constituting an utterance are dealt with as separate scenes.[4] These scenes are marked as such (*Interaction*) and are not accounted for in the analysis of the story component schema, as in (5):

(5) Healthy speaker (HP-v05)
#*не знаю*
ne zna-ju
not know-1sg.prs
'I do not know'

#*доволен он или нет*
dovolen on ili net
content-sg.m he or not
'whether he is happy or not'

#*эмоций я не замечала там особенно ни*
emotsi-j ja ne zamecha-l-a tam osobenno ni
emotion-pl.gen I.nom not notice-pst-f.sg there especially not
у кого
u kogo
at somebody.gen
'I have not noted much of emotion there'

The reason for mentioning this aspect of narrative analysis research here is that the interaction segments must be accounted for in the segmentation procedures. Sometimes they constitute a discourse unit (a clause, an utterance or even a scene) and thus may influence the quantitative results including the number of EDUs per scene.

3.2.2 *Scenes*

At the second stage of the macrolevel segmentation, the story is broken into *scenes*. At the third stage, the resulting scenes are tagged with the markers of the story genre *schema components,* and following that, EDUs constituting a scene are tagged for *components subtypes*.

The scene is a sequence of EDUs produced within one perspective. This definition is based on the concept of *stanza* by Hymes (1977), though applied not to the folklore artistic narratives, but to everyday personal stories, or test situation film retellings. We believe this transfer of the term to be logical and following our principal belief that telling stories is an art notwithstanding what kind of story it

[4] We are not discussing types of interaction markers in this paper. For a more detailed account of IMs in the CLiPS see (Bergelson & Khudyakova, 2017).

is. The artistic aspect of storytelling follows from the degree of linguistic freedom that narrators have in choosing means of verbalization for their stories. This is true for the neurologically healthy narrators, and correspondingly, our research looks into whether and to what degree this freedom of expressing oneself in the story is restricted for PWA and people with RHD.

The border between two scenes is determined by the change of the perspective. On the meta-narrative level and for the segmentation purposes, scenes are considered either part of the *world of the story*, or of the *world of narration* (Barthes, 1975; Norrick, 2000; Paducheva, 2008). Within the story world, the scenes belong to one of the following story components: *abstract, coda, description, mainline*, and *evaluation*. The abstract and the coda are, so called, fixed story components, because normally there will be only one instance of each of these scene types, and they are found at the beginning and at the end of the story, respectively. See an example of a 'multi-layer' abstract in (6):

(6) Healthy speaker (HP-v13)

#*ну я б сказала*
nu ya b skaza-l-a
well I.NOM would say-PST-SG.F
'Well, I would say'

#*не очень правдоподобная история*
ne ochen' pravdopodobn-aja istorij-a
not very credible-SG.F.NOM story-SG.NOM
'it is not a very credible story'

#*потому что*
potomu chto
because that
'because …'

#*а я должна рассказать вот например вам*
A ja dolzhn-a rasskaza-t' vot naprimer v-am
and I.NOM must-SG.F tell-INF well for.instance you.PL-DAT
'but I have to tell for instance you'

#*что вы не видели этого фильма да*
Chto v-i ne vide-l-i eto-go fil'm-a da
That you.PL-NOM not see-PST-PL this-SG.M.GEN film-SG.GEN yes
'as you have not seen the film, yeah'

#*что в нем происходит*
chto v nem proishodi-t
what in he.LOC happen-3SG.PRS
'what is going there'

#*ну*	*история*	*проста*	*как*	*мир*
nu	*istorij-a*	*prost-a*	*kak*	*mir*
well	story-SG.NOM	simple-SG.F.NOM	as	world-SG.NOM

‘well the story is as simple as it can be’

#*человек*	*падок*	*на*	*воровство*
chelovek	*pad-ok*	*na*	*vorovstv-o*
human-SG.NOM	susceptible-SG.M.NOM	on	theft-SG.ACC

‘humans are susceptible to theft’

Description, mainline, and evaluation are non-fixed: They can appear more than once at different moments in the story and often intersperse with each other. The change in perspective takes place when one or more of the following takes place:

a. Introduction of the new actor, and/or new topic, and/or new frame;
b. Shift between *story components* (e.g., from Mainline to Description);
c. Shift from one of the *fixed components* to the *non-fixed* (e.g., from Abstract to Description);
d. Shift from one meta-component to another (between any story component scene and Interaction scene).

Subtypes of the story genre schema components are tagged on the EDUs with the purpose of being used in the semantic (contents) and pragmatic analyses.

The story component *Description* includes the following subtypes: *introduction, addition, detalization,* and *clarification.* As opposed to rather similar descriptors introduced in the Rhetorical Structure Theory (Mann & Thompson, 1988), we do not use these and other story components’ subtypes for building any kind of structures, but as tags only.

Introduction mostly takes place at the beginning of the story introducing the primary scenery (tag *intro*), as in (7):

(7) Healthy speaker (HP-v05)

#*показывают*	*сельскую*	*местность*
pokazyvaj-ut	*sel’sk-uju*	*mestnost*
show-3SG.PRS	rural-SG.F.ACC	area-SG.ACC‘

‘(They) show a rural area’

Addition marks EDUs introducing information which is new at this point of discourse (tag *add*). Typically, additions are accompanied by discourse markers *pri etom* ‘so’ and *takzhe* ‘also, and’.

The *detalization* subtype name speaks for itself. It provides more details of the already introduced information (tag *det*). *Clarification* is different from detalization in pointing to the character of the relation between the new information and the story events – causal, temporal, conditional, or restrictive relations.

The *Mainline* component – the narrative nucleus of any story depicting events that take place in the story world – embraces two subtypes: events and quasi-events. *Events* are EDUs that render actions and processes – real events in the physical world of the story (tag *pred*). By *quasi-events* we understand EDUs with epistemic predicates and locutionary verbs that are very frequent in the context of our stories (retellings of the film) and express quasi-speech, quasi-thought and quasi-perception of the characters as guessed by the narrator (tag *qpred*). The element "quasi" used to characterize these predicates points to the specifics of our experimental design. While retelling the film they have watched, our narrators try to express in their stories the mental activities of the participants of the narrated events. One can see what a character in the film is doing (like picking pears or dragging the goat) but not what the character thinks, decides and believes. Discriminating between these two classes of events may be important when we compare various narrative strategies, especially for the PWA and RHD: the cognitive load to verbalize events that took place on the screen *versus* quasi-events that one has to guess about, may differ significantly.

The content of speech or thought introduced by quasi-predicates is not part of the story mainline and constitutes another component in the story world – *Evaluation*. It consists of opinions and judgments of the story characters as guessed by the narrator (*mal'chik reshil, chto grushi nich'i* 'the boy decided that the pears don't belong to anyone'; *sadovnik podumal, chto eto ego grushi* 'the gardener thought they were his pears').

The subtype *Content* serves for introducing the speech or thoughts of the story characters in an indirect manner (tags *det, est* or *jdg*), while *Citation* does the same but directly, using deictic categories demonstrating empathy of the narrator with the character whose speech or thoughts are being rendered (tag *cit*). The tag *est* is used for assessments of the event probability and the tag *jdg* – for the opinions in terms of "liking ~ disliking", or "good ~ bad". All of it – from the perspective of the story character –, while evaluations and assessments made by the narrator, belong to the narration world and are part of the interaction with the addressee. As such, the latter are marked on a separate annotation layer saved for markers of interaction.

3.2.3 *Segmentation criteria*

Thus, the segmentation criteria used at the macrolevel are semantic and/or pragmatic, supported by some lexical and grammatical markers. On the other hand, the intonation contour for larger chunks of discourse (final falling intonation) can serve as a typical signal of completeness either on informational (event/situation completeness), or on interactional (end of turn) levels. It is also accompanied by significant pauses, as compared to the average pause for this speaker. Below in (8) we demonstrate one story by an aphasia speaker, broken in scenes tagged for the story component analysis. Additionally to the regular segmentation process at that level, we mark the final intonation contour (falling or raising), longer pauses (over **0.6** seconds) and some other information. The story is produced by a person with a fluent aphasia. Failures to produce correct nominations are underlined in the English translation line. Grammatical errors like wrong case or gender inflexions which go unnoticed by the narrator are underlined in the transliterated text and the glosses line.

(8) Speaker with sensory aphasia (AP-s07)

#1. (Interaction)

#*что*	(0.2)	*можно*	*начинать*	*теперь?*	(**2.1**)
chto	(0.2)	*mozhno*	*nachinat'*	*teper'?*	(**2.1**)
what		possible	start-INF	now	

'so can I start now?'

#2. (Abstract)

#*на*	*этой*	э	(0.5)	*с*	(0.9)	*фильме*	*показано*
na	*etoj*	*e*	(0.5)	*s*	(0.5)	*fil'me*	*pokaza-n-o*
on	this-SG.F.LOC					film-SG.LOC	show-PTCP-SG.N.NOM

оборот	(0.5)	*у*	(0.6)	оборот	э	(0.7)	(1.5)	а	(1.0)	э	ру=
oborot	(0.5)	*u*	(0.6)	oborot	e	(0.7)	(1.5)	a	(1.0)	eh	ru=
turnover-SG.ACC				turnover-SG.ACC							

у=	*фру=*	*фруктов*	*на*	*одном*	*из*	э	э	*сада*	(**0.7**)
u=	*fru=*	*frukt-ov*	*na*	*odn-om*	*iz*	*e*	*e*	*sad-a*	(**0.7**)
		fruit-PL.GEN	on	one-SG.M.LOC	from			garden-SG.GEN	

'this film demonstrates how fruits are picked in one garden'

#3. (intro)

#*молодой*	*э*	*м*	*мужик*	*муж=*	*мужчина*	*с=*	*со=*
molod-oj	*e*	*m*	*muzhik*	*muzh=*	*muzhchin-a*	*s=*	*so=*
young-SG.M.NOM			man-SG.NOM		man-SG.NOM		

собрывал	*э*	*груши*
sobryv-a-l	*e*	*grush-i*
gather-PST-SG.M		pear-PL.ACC

'young man was picking pears'

#*и*	*собирал*	*их*	*в*	*большие*	*коро=*	*корзины*	**(1.0)**
i	*sobira-l*	*ih*	*v*	*bol’sh-ie*	*koro=*	*korzin-y*	**(1.0)**
and	gather-PST-SG.M	they.ACC	in	big-PL.ACC		basket-PL.ACC	

‘and was putting them in large baskets’

#4. (pred)

#*в*	*это*	*время*	*прошел*	*мужчина*	*с*
v	*et-o*	*vremj-a*	*proshe-l*	*muzhchin-a*	*s*
in	this-SG.N.ACC	time-SG.ACC	pass-PST-SG.M	man-SG.NOM	with

козой	**(1.2)**
koz-oj	**(1.2)**
goat-SG.INSTR	

‘at this moment a man with a goat passed by’

#5. (pred)

#*а*	*потом*	*пробежал*	*мальчик*	**(1.1)**
a	*potom*	*probezha-l*	*mal’chik*	**(1.1)**
and	then	run-PST-SG.M	boy-SG.NOM	

‘and then a boy ran by’

#6. (Interaction)

#*а*	*не*	*знаю*
a	*ne*	*znaj-u*
and	not	know-1SG.PRS

‘I don’t know’

#*о*	*чем*	*там*	*разговаривал*
o	*ch-em*	*tam*	*razgovariva-l*
about	what-LOC	there	talk-PST-SG.M

‘what he was talking about’

#7. (pred)

#*но*	*тем*	*не*	*менее*	*он*	*забрал*	*одну*	*из*
no	*tem*	*ne*	*menee*	*on*	*zabra-l*	*odn-u*	*iz*
but	that.INSTR	not	little.COMP	he.NOM	take-PST-SG.M	one-SG.F.ACC	from

э=	*этих*	*а*	*б=*	*т=*	*к=*	*корзину*	*с*	*этими*	*э*	*с*
e=	*et-ih*	*a*	*b=*	*t=*	*k=*	*korzin-u*	*s*	*et-imi*	*e*	*s*
	this-PL.GEN	and				basket-SG.ACC	with	this-PL.GEN		with

грушами	*и* (0.5)
grush-ami	*i* (0.5)
pear-PL.INSTR	

‘but anyway, he took one of these baskets with these pears’

#*повез*	*на*	*велосипеде*
povez	*na*	*velosiped-e*
carry-PST.SG.M	on	bike-SG.LOC

‘and went on a bike’

#8. (pred)
#*во время* (0.8) *э в это время встретился э*
vo vremj-a (0.8) *e v et-o vremj-a vstreti-l-sja e*
in time-SG.ACC in this-SG.N.ACC time-SG.ACC meet-PST-SG.M
девчонка ему
devchonk-a emu
girl-SG.NOM he.ACC
‘this time, at this time he came across a girl’
#*он* (**0.8**) *э нечаянно зацепился*
on (0.8) *e nechajanno zacepi-l-sja*
he.NOM accidentally trip-PST-SG.M
‘he accidentally tripped’
#*и упал э* (**0.6**) (**0.8**) *у* (**0.6**)
i upa-l e (0.6) (0.8) u (0.6)
and fall-PST-SG.M
‘and fell’

#9. (pred)
#*была упала эта самая гру= э*
by-l-a upa-l-a et-a samaj-a gru= e
be-PST-SG.F fall-PST-SG.F this-SG.F.NOM exact-SG.F.NOM
корзина с э м э грушами (**1.1**)
korzin-a s e m e grush-ami (1.1)
basket-SG.NOM with pear-PL.INSTR
‘this very basket with pears fell (down)’

#10. (pred)
#*а мимо проезжали ребя= про= проходили ребята*
a mimo proezzha-l-i rebja= pro= *prohodi-l-i rebjata*
and by ride-PST-PL walk.by-PST-PL kid.PL.NOM
‘and kids were passing by’
#*они спо= помогли ему собрать эти э груши* (**0.9**)
oni spo= pomog-l-i emu sobra-t’ et-i e grush-i (**0.9**)
they.NOM help-PST-PL he.DAT pick-INF this-PL pear-PL.ACC
‘they helped him to pick these pears’

#11. (pred)
#*и э по полю пошли* (**0.5**)
i e po pol-ju posh-l-i (**0.5**)
and on field-SG.DAT walk-PST-PL
‘and they started walking along the field’

#12. (pred)

#*но ребята нашли его* (2.1) э *сла= шляпу* (0.8)
no rebjat-a nash-l-i ego (2.1) e *sla= shljap-u* (0.8)
but kid.PL-NOM find-PST-PL his hat-SG.ACC
‘but the kids found his hat’

#*и ему вернули*
i emu vernu-l-i
and he.ACC return-PST-PL
‘and gave it back to him’

#*он как бы за вместо этого дал им*
on kak by za vmesto et-ogo da-l im
he.NOM as.if for instead this-SG.GEN give-PST-SG.M they.DAT
подарок три грушки (**1.2**)
podarok tri grushk-i (**1.2**)
present.ACC three pear-SG.GEN
‘He gave them three pears sort of a present instead of it’

#13. (pred)

#*и они пошли мимо этого сра= сада*
i oni posh-l-I mimo et-ogo sra= sad-a
and they-NOM go-PST-PL by this-SG.M.GEN garden-SG.GEN
мимо (**0.6**)
mimo (**0.6**)
by
‘And they went along this garden’

#14. (qpred)

#*а мужчи= м товарищ*
a muzhchi= m tovarishch
and comrade-SG.NOM
‘and the guy…’

#*который убирал эти груши*
kotor-yj ubira-l et-I grush-i
which-SG.M.NOM pick-PST-M this-PL.ACC pear-PL.ACC
‘who was picking those pears’

#*не мог понять* (**0.6**)
ne mog ponja-t’ (**0.6**)
not can.PST-M understand-INF
‘…could not make sense’

#15. (det)

#*во-первых*	*пропала*	*как.бы*	(0.7)	*одна*	*корзина*
vo-pervyh	*propa-l-a*	*kak.by*	(0.7)	*odn-a*	*korzin-a*
firstly	disappear-PST-SG.F	sort.of		one-SG.F.NOM	basket-SG.NOM

‘firstly, one basket sort of got lost’

#*которую*	*он*	*чёто*	*не*	*не*	*поймет*
kotor-uju	*on*	*chjoto*	*ne*	*ne*	*pojm-et*
which-SG.F.ACC	he	something	not	not	understand-PRS.3SG

‘that he does not understand (it)’

#*и*	*проходят*	*ребята*
i	*prohodj-at*	*rebjat-a*
and	pass.by-PRS.3PL	guys-NOM

‘And the kids pass by’

#*и*	*кушают*	*его*	*их*	*груши*	**(1.3)**
i	*kushaj-ut*	*ego*	*ih*	*grush-I*	**(1.3)**
and	eat-PRS.3PL	his	their	pear-PL.ACC	

‘and eat his pears’

#16. (Coda)

#*ну*	*и*	*на*	*этом*	*все*	*закончилось*
nu	*i*	*na*	*et-om*	*vse*	*zakonchi-l-o-s’*
well	and	on	this-ACC	all-NOM	end-PST-SG.N

‘so this is how it ended’

3.2.4 *Analysis*

The story in (8) illustrates various combinations of the prosodic (final intonation contour and longer pauses at the assumed scene borders) and semantic and pragmatic features that together allow for the scenes segmentation. All the scenes end with a falling intonation contour except the scene #1 tagged as *Interaction*, which is a question addressed to the experimenter. In the majority of cases (14 out of 16 scenes), the “default” combination of prosodic and pragma-semantic features takes place: Final prosodic contour (falling intonation) combines with a pause over 0.6 s and corresponds to the change in the storytelling perspective:

1. Shift between interaction and the world of the story:
   a. #1 → #2 – Interaction → Abstract
   b. #5 → #6 – Mainline → Interaction
2. Shift between the story components:
   a. #2 → #3 – Abstract → Description: *intro*
   b. #3 → #4 – Description → Mainline: *pred* (see Figure 4)
   c. #14 → #15 – Mainline → Description: *det*
   d. #15 → #16 – Description → Coda

3. Shift between the actors, or the time and/or place:
   a. #4 → #5 – Mainline → Mainline: *pred*
   b. #9 → #10 – Mainline → Mainline: *pred*
   c. #10 → #11 – Mainline → Mainline: *pred*
   d. #11 → #12 – Mainline → Mainline: *pred*
   e. #12 → #13 – Mainline → Mainline: *pred*
   f. #13 → #14 – Mainline → Mainline: *qpred*

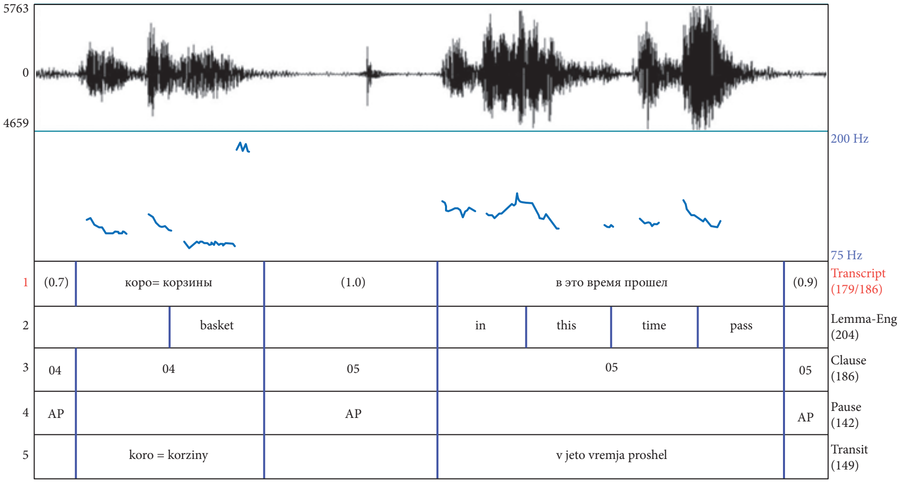


Figure 4. From scene #3 →to scene #4: Pause and prosodic contour

Of the two scenes starting without a longer pause at the beginning, one (#7) immediately follows the scene tagged as interaction, thus resuming the story line interrupted by #6.

The other scene without a longer pause at the beginning (#8) has a long pause after the first NP (*vo vremja* 'at the time'), which is a false start immediately followed by a repair. Such false starts at the beginning of a new scene that involve anaphoric pronouns or formulaic expressions (in the next utterance of the same scene #8), correspond well with the idea of production planning while turn taking (Holler, Kendrick, Casillas, & Levinson, 2015). It must be mentioned though, that in case of the clinical discourse it is often impossible to discriminate between the situations of production planning and word retrieval difficulties.

Segmentation even for the larger units like scenes cannot rely on the prosodic parameters only. We observe similar structures, where presence or absence of the final falling tone and longer pause after it cannot determine the border between the scenes – compare #8 with #12. Both scenes contain two EDUs united by the main character (the boy) with symmetric pronominal reference (*emu* || *on* 'him || he') and syntactic structure.

Compare in #8 … *vstretilsja [e] devchonka* ***emu*** **||** ***on*** (0.8) *[e] nechajanno zacepilsja* … '… a girl came across him || he accidentally tripped …' and in #12 … (0.8) *i* ***emu*** *vernuli* || ***on*** *kak by* *<u>za vmesto etogo</u>* *dal im podarok tri grushki* '… and gave it back to him || he gave them three pears sort of a present <u>instead of it</u>'

However, in #8 two EDUs within the scene are separated with the final prosodic contour and a long pause, while in #12 none is present. The story in (8) contains a long and uninterrupted sequence of the mainline scenes. It is typical for the stories by PWA to reduce the cognitive load of storytelling by maximally sticking to mainline only and not getting distracted by evaluations and descriptions as much as possible. The same reasoning also explains the wider range of the number of clauses per scene for PWA and RHD as compared to the healthy participants. The average length of a scene for the latter is 4–6 clauses per scene. For the PWA and RHD participants it is 2–7 clauses per scene. The problem with this measure is that scenes that are tagged as *interaction* are excluded from the count, but interaction clauses that are part of other scene types do count in the number of clauses.

4. Conclusion

In this chapter we did not aim at a detailed analysis of various peculiarities of the clinical discourse samples; rather our goal was to demonstrate problems that emerge when well-established and robust procedures of discourse segmentation, using acoustic and prosodic features, are applied to the clinical discourse, as represented by stories based on visual stimuli. Though the CLiPS corpus is quite unique due to the combination of its features (Russian language clinical discourse based on the "Pear Stories" film) the problems with segmentation aiming at basic units are not unique for the clinical linguistics at all. Being unable to rely on uniform prosodic segmentation criteria, we opted for the clause as a broadly understood EDU. We compensate for the absence of reliable acoustic basis for segmentation by using this very information in the annotation layers: absolute and filled pauses measured in milliseconds, prosodic contours, repetitions and false starts. Creating this corpus, we aim at getting a resource for the multilevel and multifaceted analysis of the clinical discourse. Intergroup comparison of different types of aphasia and RHD involves many different issues in discourse production and comprehension. To research them, we operate with three types of discourse units: clauses (as EDUs), utterances and scenes. In combination, they allow us to study various features and parameters both on the micro- and macro-level of discourse analysis, including fluency, syntactic complexity, informativeness, coherence, empathy, interaction markers, and story genre schema.

Acknowledgements

The book chapter is based on the study funded by the Center for Language and Brain NRU Higher School of Economics, RF Government grant, ag. №14.641.31.0004.

References

Akhutina, T. (2015). Luria's classification of aphasias and its theoretical basis. *Aphasiology*, 30(8), 878–897. https://doi.org/10.1080/02687038.2015.1070950

Alexander, M. P., & Hillis, A. E. (2008). Aphasia. *Handbook of Clinical Neurology*, 88, 287–309. https://doi.org/10.1016/S0072-9752(07)88014-6

Armstrong, E., Ciccone, N., Godecke, E., & Kok, B. (2011). Monologues and dialogues in aphasia: Some initial comparisons. *Aphasiology*, 25(11), 1347–1371. https://doi.org/10.1080/02687038.2011.577204

Bamberg, M. (2012). Narrative analysis. In *APA handbook of research methods in psychology: Vol 2, Research designs: Quantitative, qualitative, neuropsychological, and biological*. (pp. 85–102). Washington DC: American Psychological Association. https://doi.org/10.1037/13620-006

Barthes, R. (1975). On narrative and narratives: An introduction to the structural analysis of narrative. *New Literary History*, 6, 237–272. https://doi.org/10.2307/468419

Bergelson, M. B. (2007). Sociokul'turnaja motivirovannost' narrativov: analiz lichnyh rasskazov [Socio-cultural motivation of narratives: Analysis of personal stories]. In M. Bergelson (Ed.), *Pragmaticheskaja i sotsiokulturnaja motivirovannost' jazykovoj formy*. [Pragmatic and sociocultural motivation of linguistic form] (pp. 69–111). Moscow: Universitetskaya kniga.

Bergelson, M., & Khudyakova, M. (2017). Interaction and empathy as elements of narrative strategies in the Russian CliPS corpus. *Computational Linguistics and Intellectual Technologies* (Vol. 2), 16(23), 55–67.

Berman, R. (1997). Narrative theory and narrative development: The Labovian impact. In *Oral versions of personal experience: Three decades of narrative analysis*. Special issue of *Journal of Narrative and Life History*, 7(1–4), 235–244. Amsterdam: John Benjamins.

Boersma, P., & Weenink, D. (2007). *Praat: Doing phonetics by computer* (Version 4.5.) [Computer program]. <http://www.praat.org/>

Carlson, L., & Marcu, D. (2001). Discourse tagging reference manual. *ISI Technical Report ISI-TR-545*, 2, 1–87.

Chafe, W. (Ed.). (1980). *The Pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.

Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago, IL: University of Chicago Press.

Chafe, W. (2008). The analysis of discourse flow. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 671–687). Oxford: Blackwell. https://doi.org/10.1002/9780470753460.ch35

Chafe, W. (2014). Language and the flow of thought. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (Vol.1, pp. 93–112). Mahwah, NJ: Lawrence Erlbaum Associates.

Coelho, C. A. (2002). Story narratives of adults with closed head injury and non-brain-injured adults: Influence of socioeconomic status, elicitation task, and executive functioning. *Journal of Speech, Language, and Hearing Research*, 45(6), 1232–1248. https://doi.org/10.1044/1092-4388(2002/099)

De Fina, A., & Georgakopoulou, A. (2012). *Analyzing narratives*. Cambridge: Cambridge University Press.

De Fina, A., & Georgakopoulou, A. (2015). *The handbook of narrative analysis*. Hoboken, NJ: Wiley-Blackwell. https://doi.org/10.1002/9781118458204

Degand, L., & Simon, A. C. (2005). Minimal discourse units: Can we define them, and why should we. *Proceedings of SEM-05. Connectors, Discourse Framing and Discourse Structure: From Corpus-Based and Experimental Analyses to Discourse Theories*, 477, 65–74.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language : A unit for all reasons. *Applied Linguistics*, 21(3), 354–375. https://doi.org/10.1093/applin/21.3.354

Glosser, G., & Deser, T. (1991). Patterns of discourse production among neurological patients with fluent language disorders. *Brain and Language*, 40(1), 67–88. https://doi.org/10.1016/0093-934X(91)90117-J

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *Boston diagnostic aphasia examination: Short from record booklet*. Philadelphia, PA: Lippincott Williams & Wilkins.

Heilman, K. M., Leon, S. A., & Rosenbek, J. C. (2004). Affective aprosodia from a medial frontal stroke. *Brain and Language*, 89(3), 411–416. https://doi.org/10.1016/j.bandl.2004.01.006

Holland, A. L. (1980). *CADL communicative abilities in daily living: A test of functional communication for aphasic adults*. Baltimore, MD: University Park Press.

Holland, A. L. (1982). Observing functional communication of aphasic adults. *Journal of Speech and Hearing Disorders*, 47(1), 50–56. https://doi.org/10.1044/jshd.4701.50

Holler, J., Kendrick, K. H., Casillas, M., & Levinson, S. C. (2015). Editorial: Turn-taking in human communicative interaction. *Frontiers in Psychology*, 6, 1919. https://doi.org/10.3389/fpsyg.2015.01919

Hymes, D. H. (1977). Discovering oral performance and measured verse in American Indian narrative. *New Literary History*, 8, 431–457. https://doi.org/10.2307/468294

Johnstone, B. (2016). Oral versions of personal experience: Labovian narrative analysis and its uptake. *Journal of Sociolinguistics*, 20(4), 542–560. https://doi.org/10.1111/josl.12192

Khudyakova, M. V., Bergelson, M. B., Akinina, Y. S., Iskra, E. V., Toldova, S., & Dragoy, O. V. (2016). Russian CliPS: A corpus of narratives by brain-damaged individuals. In D. Kokkinakis (Ed.), *Proceedings of LREC 2016 workshop* (pp. 22–26). Portorož, Slovenia: Linköping University Electronic Press.

Kibrik, A. A., & Podlesskaya, V. I. (Eds.). (2009). *Rasskazy o snovidenijah: korpusnoe issledovanie ustnogo russkogo diskursa* [Night dream stories: A corpus study of spoken Russian discourse]. Moscow: Languages of Slavonic Culture.

Labov, W. (1972). The transformation of experience in narrative syntax. In W. Labov (Ed.), *Language in the inner city: Studies in the black English vernacular* (pp. 354–396). Philadelphia, PA: University of Pennsylvania Press.

Labov, W. (1997). Some further steps in narrative analysis. In *Oral versions of personal experience: Three decades of narrative analysis. Special issue of Journal of Narrative and Life History*, 7(1–4), 395–415. Amsterdam: John Benjamins.

Labov, W., & Waletzky, J. (1967). Narrative analysis. In J. Helm (Ed.), *Essays on the verbal and visual arts* (pp. 12–44). Seattle, WA: University of Washington Press.

Linnik, A., Bastiaanse, R., & Höhle, B. (2015). Discourse production in aphasia : A current review of theoretical and methodological challenges. *Aphasiology*, 30(7), 765–800. https://doi.org/10.1080/02687038.2015.1113489

Linnik, A. S., Bastiaanse, R., & Khudyakova, M. V. (2015). What contributes to discourse coherence? Evidence from Russian speakers with and without aphasia. *Stem-, Spraak- En Taalpathologie*, 20(1), 107–110.

Luria, A. R. (1972). Aphasia reconsidered. *Cortex*, 8(1), 34–40. https://doi.org/10.1016/S0010-9452(72)80025-X

MacWhinney, B. (2010). Part 1: The CHAT Transcription format. *The CHILDES Project: Tools for Analyzing Talk*. Mahwah, NJ: Lawrence Erlbaum Associates. https://doi.org/10.1111/1460-6984.12101/abstract

Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text – Interdisciplinary Journal for the Study of Discourse*, 8(3), 243–281. https://doi.org/10.1515/text.1.1988.8.3.243

Marini, A., Andreetta, S., Tin, S., Carlomagno, S., del Tin, S., & Carlomagno, S. (2011). A multilevel approach to the analysis of narrative language in aphasia. *Aphasiology*, 25(11), 1372–1392. https://doi.org/10.1080/02687038.2011.584690

Marini, A., Galetto, V., Zampieri, E., Vorano, L., Zettin, M., & Carlomagno, S. (2011). Narrative language in traumatic brain injury. *Neuropsychologia*, 49(10), 2904–10. https://doi.org/10.1016/j.neuropsychologia.2011.06.017

Meuse, S., & Marquardt, T. P. (1985). Communicative effectiveness in Broca's aphasia. *Journal of Communication Disorders*, 18(1), 21–34. https://doi.org/10.1016/0021-9924(85)90011-5

Miller, J. F., Andriacchi, K., & Nockerts, A. (Eds.). (2015). *Assessimg language production using SALT software. A clinician's guide to language sample analysis*. Middleton, WI: SALT Software LLC.

Norrick, N. (2000). *Conversational narrative: Storytelling in everyday talk*. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.203

Paducheva, E. V. (2008). Diskursivnye slova i kategorii: rezhimy interpretacii [Discourse words and categories: Interpretation regimes]. In V. Plungyan (Ed.), *Issledovaniya po teorii grammatiki* [Studies in grammar theory] (Vol. 4, pp. 56–86). Moscow: Gnozis.

Passonneau, R. J., & Litman, D. J. (1997). Discourse segmentation by human and automated means. *Computational Linguistics*, 23(1), 103–139.

Polanyi, L. (1985). A theory of discourse structure and discourse coherence. *Papers from the General Session at the Twenty-First Regional Meeting*, 21(1), 306–322.

Polanyi, L. (1989). *Telling the American story. A structural and cultural analysis of conversational storytelling*. Cambridge, MA: The MIT Press.

Raso, T., & Mello, H. (2014). *Spoken corpora and linguistic studies*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61

Seddoh, A. S. (2004). Prosodic disturbance in aphasia: Speech timing versus intonation production. *Clinical Linguistics & Phonetics*, 18(1), 17–38. https://doi.org/10.1080/0269920031000134686

Shewan, C. M. (1988). The Shewan Spontaneous Language Analysis (SSLA) system for aphasic adults: Description, reliability, and validity. *Journal of Communication Disorders*, 21(2), 103–138. https://doi.org/10.1016/0021-9924(88)90001-9

Speech Analyzer – SIL Language Technology (Version 3.1) [Computer software]. Retrieved from <https://software.sil.org/speech-analyzer/#about>

Swinburn, K., Porter, G., & Howard, D. (2004). *CAT: Comprehensive aphasia test*. Hove: Psychology Press.

Taboada, M., & Zabala, L. H. (2008). Deciding on units of analysis within Centering Theory. *Corpus Linguistics and Linguistic Theory*, 4(1), 63–108. https://doi.org/10.1515/CLLT.2008.003

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In *Proceedings of LREC 2006* (pp. 1556–1559). Genoa, Italy.

CHAPTER 9

Cross-linguistic comparison of automatic detection of speech breaks in read and narrated speech in four languages

Plínio A. Barbosa
State University of Campinas, Institute for Language Studies, CNPq

This chapter tests an algorithm for the automatic detection of speech breaks in read and narrated speech in Brazilian Portuguese (BP), European Portuguese (EP), French, and German. The algorithm is independent of previous transcription or linguistic analysis (syllable, phone labeling and segmentation), requiring only the audio file. It operates in two stages: vowel onsets detection firstly, followed by V-to-V duration intervals normalization for smoothed duration z-scores. Peaks over 2.5 of the latter were considered speech breaks. Compared to human segmentation, hits for reading (70%) were higher than for narration (60%). Crosslinguistic results show EP and French having the highest proportion of hits. A test with the English *Navy* audio file reveals a hit proportion similar to German.

Keywords: automatic speech segmentation, duration, prosodic boundary, cross-linguistic comparison

1. Introduction

With the availability of huge corpora referred to "big data", it has become crucial to ensure the automatic segmentation of appropriate linguistic units for several levels of analysis, including those of semantic, syntactic, phonological and phonetic kind. In this regard, the less controversial units of segmentation in speech research are prosodic constituents referred to as intonation phrases (also as intonation units) and intonation utterances. These boundaries are respectively called as non-terminal and terminal here and in part of the literature (Botinis, Granström, & Möbius, 2001; Cresti, 2000).

The perception of such boundaries or breaks is associated to the classical prosodic parameters of fundamental frequency (f0), duration and intensity. In fact, a recent review of acoustic correlates of boundary marking in spontaneous speech

https://doi.org/10.1075/scl.94.09bar
Additional files available at http://doi.org/10.1075/scl.94.09bar/audio

in languages such as English, French and Brazilian Portuguese revealed that the most systematic cues for the perception of a terminal or non-terminal break are the presence of a silent pause, a lengthening of the pre-boundary syllable, a rise or fall in fundamental frequency (f0) contour and a fall in intensity as a covariant parameter mostly associated to terminality (Mittman & Barbosa, 2016). Less often, but consistently, cross-boundary intensity change, f0 reset and pre-boundary creaky voice have a relevant role in signaling boundary as well.

The study of the order of importance of the boundary-related acoustic parameters in a particular context is a crucial step for the building of an algorithm for automatically detecting boundaries in speech. In this regard, previous research in English (Campbell, 1993; Ni, Zhang, Liu, & Xu, 2012; Wightman et al., 1992), French (Barbosa, 1994), Mandarin (Ni et al., 2012) and Brazilian Portuguese (Barbosa, 1996, 2007) revealed that syllable-sized duration lengthening and pausing are the leading parameters for boundary marking. More comments on that issue are made below.

The best prediction models for English sentence and phrase boundaries in spontaneous speech exhibit performances that go from 74% to 93%. The work by Ni et al. (2012) evaluated the performance of neural network and decision tree classifiers for predicting prosodic break (utterance boundary) in both Mandarin Chinese and American English. They associated syllable-based acoustic prosodic (with a neural network) and lexical and syntactic features (with a decision tree) to predict breaks (Mandarin and English) and intonational phrases (English). For the sake of comparison, only the predictions from acoustic features are discussed here. From Mandarin, the authors extracted the following parameters: three duration-related syllable features, one silent pause duration feature, 15 f0-related features, and 11 energy-related features to achieve a maximum of 85% of precision in detecting a break. As for English, 10 f0-related, 10 energy-related, and four duration-related features were combined with lexical and syntactic features to achieve 82% of precision in predicting breaks. Duration-related features combined were by far the most relevant parameters for predicting breaks in both languages with syllable and pause duration as the top predictors.

The first stage of the system developed by Kim (2004) predicts sentence boundaries and locations of interruption in speech from decision trees trained with prosodic and lexical features and a language model. Their corpus is a subset of the Switchboard corpus (Godfrey, Holliman, & McDaniel, 1992) of conversational telephone speech from several dialects of American English. The prosodic features used were duration-, f0- and energy-related features as well as speaker turn features. Duration-related included word and rhyme durations, rhyme duration differences between two neighboring words, and silence duration following a word. F0-related features were obtained from a smoothed f0 contour and included minimum, mean, and maximum value of f0 contour over a word, slope of f0 contour at the end of a word, and differences in f0 statistics and f0 values at the start and end of two neighboring words at the candidate boundary. Energy features were minimum, mean and maximum energy value of a word and its rhyme. Both f0 and energy features were normalized by mean f0 and energy values calculated over each conversation side. The author included eight speaker turn features for dealing with turns in the conversations across the telephone line. Lexical features consisted of part-of-speech information as well as information on the existence of filler words to the right of the current word boundary. With the use of prosody only for prediction, sentence boundary detection precision was 74% and Intonational Phrase (IP) boundary detection precision was 56%.[1] The addition of lexical information raises precision of detection respectively to 79% and 77%, due mainly to the possibility of distinguishing grammatical words from lexical words and fillers. For the sake of our work, it is important to investigate the relative importance of the acoustic parameters for boundary prediction, as evaluated in the following paper.

Shriberg, Stolcke, Hakkani-Tür, and Tür (2000) indicated the same results more recently obtained by Kim (2004), namely that prosodic information is enough to achieve more than 70% of sentence boundary detection.[2] They worked with American English with the Broadcast news corpus and the Switchboard corpus and used an extended set of prosodic and speaker turn features similar to the ones used by Kim, but extracted them within a narrow window of 200 ms before and after the sentence boundary potential location. For Switchboard, the best performance revealed this hierarchy of relevant parameters for utterance boundary detection: phone and rhyme duration preceding boundary (49%), pause duration at boundary (18%), turn/no turn at boundary (17%), and pause duration at previous word boundary (15%). Performance achievement for that case was 93%.

Also Gotoh and Renals (2000) have shown that it is possible to detect utterance boundaries in a corpus of British English broadcast news from BBC 1 with

[1] Sentence units (SUs) were defined by the author as "linguistic units roughly equivalent to sentences that are used to mark segmentation boundaries in conversational speech where utterances are often completed without forming 'grammatical' sentences in the sense one thinks of with written text" (Kim, 2004, p. 4). They subcategorized them in statements, questions, backchannels, and incomplete SUs. The first two types were considered SUs if they form grammatically complete sentences, phrases or clauses that functions as a standalone entity, whereas "backchannels refer to words or phrases used during a conversation to provide feedback or acknowledgement to the dominant speaker, indicating that the conversation partner is listening" (Kim, 2004, pp. 4–5). Incomplete SUs occurred "when the speaker trailed off and abandoned the SU without completing it or when the speaker was interrupted by another speaker before the SU could be finished" (Kim, 2004, p. 5).

[2] In their work, sentence boundaries were automatically determined by using a tagger that was trained on the basis of a segmentation derived from the capitalization and punctuation in the corresponding transcripts.

a precision rate of 74% with pause duration alone.[3] These boundaries are mainly associated with terminal boundaries when their results are closely examined.

The works reviewed so far suggest that duration-related features are the most relevant parameters for predicting prosodic boundary. In fact, Campbell (1993) has previously shown that differences of normalized segmental duration in a syllable frame allow the appropriate prediction and distinction of stress-related from boundary-related segmental lengthening in British English isolated read sentences. Based on this investigation, he developed an algorithm for detecting phrase boundaries, which achieved an agreement of 73% for boundary insertion with four speakers in 200 isolated sentences.

The algorithms used in the works just reviewed require training. That is, all these algorithms learn how to associate the acoustic data with the boundaries by adapting weights from their formulae to classify the boundaries (no boundary/boundary). To test these algorithms with different languages would require a specific training phase. Although developed for French in order to detect period boundaries and prominences, the ANALOR software (Avanzi, Lacheret, & Victorri, 2008; Lacheret-Dujour, Simon, Goldman, & Avanzi, 2013) can be used for tracking intonational units (IUs), as was done by Mettouchi, Lacheret-Dujour, Silber-Varod, and Izre'el (2007) for Kabyle and Hebrew. In comparison with human perception by a native expert, the automatic hit rate for IUs, which includes terminal and non-terminal boundaries, was 100% for Kabyle and 71% for Hebrew by using narratives. A similar figure as the one for Hebrew was obtained with narration in the work we present here. One disadvantage of the algorithm used by ANALOR in comparison with the one shown here is that it requires a previous syllabic segmentation for predicting boundaries.

The performance of the algorithms presented so far for English and Mandarin and the prevalence of duration-related features for utterance boundary detection stimulated us to fully investigate how a higher performance can be achieved by using a more precise use of prosodic duration, namely, syllable-sized duration. And because performance of speech break detection can depend on language and style, we decided to test our algorithm with four languages (French, German, European and Brazilian Portuguese) in two styles (reading and narration).

In Section 2 we present the corpus and describe the algorithm implemented as a Praat (Boerma & Weenink, 2017) script. In Section 3 we present the results for non-terminal and terminal break prediction. In Section 4 we discuss the results obtained and propose some directions for increasing break prediction performance in speech.

[3] The BBC corpus does not include weather reports, but no other information on the kind of news is given.

2. Methodology

2.1 Corpus

The CROSS-RHYTHM corpus consists of parallel productions in Brazilian Portuguese (BP),[4] European Portuguese (EP),[5] standard French (FR),[6] and standard German (GE).[7] It is formed by reading and storytelling materials from ten subjects (5 males and 5 females) in each language or variety. The text read by the ten subjects was "The Awkward Monk", a circa 1,600-word text originally written in EP about the origin of the Belém pastries. The text was adapted to BP and translated sentence by sentence into French and German. The same speakers immediately told the story they had just read in their own languages. All subjects of BP, French, and German were university graduates aged between 25 and 40 at the time of recordings, whereas all EP speakers were full researchers of INESC-Lisboa, Portugal around 50 years of age. It is important to say that EP speakers associated information found in the story they read with their own experience with the pastries, since Belém is a well known place in Lisbon. Subjects do not have professions in which they use their voice professionally.

From this corpus we selected two female speakers in each language in the two speaking styles in order to test the proposed algorithm. There is no particular reason for choosing female speakers, besides the fact that they produced longer narratives. All terminal boundaries were associated to sentence boundaries in the reading style for reasons of comparison across languages. As for the storytelling style, each terminal boundary was determined from perception grounds by relating each utterance to a communicative act (Cresti, 2000). Non-terminal breaks were associated in both styles to intonational phrases corresponding to incomplete communicative acts.

2.2 The SalienceDetector script

For detecting duration-related acoustic salience in large corpora without the need of any previous segmentation or labeling, the *SalienceDetector* script for Praat was conceived in two stages. The first stage automatically detects vowel onsets by the

[4] Brazilian Portuguese audios analyzed: ▶ bp_np_nr, ▶ bp_np_re, ▶ bp_ra_nr, ▶ bp_ra_re.

[5] European Portuguese audios analyzed: ▶ ep_am_nr, ▶ ep_am_re, ▶ ep_ar_nr, ▶ ep_ar_re.

[6] French audios analyzed: ▶ fr_ca_nr, ▶ fr_ca_re, ▶ fr_ma_nr, ▶ fr_ma_re.

[7] German audios analyzed: ▶ ge_s5_nr, ▶ ge_s5_re, ▶ ge_s6_nr, ▶ ge_s6_re.

use of the Beat Extractor algorithm developed by Cummins and Port (1998). This algorithm was modified to include two filtering techniques as its first stage and additional criteria for vowel onset detection from acoustics, as shown in step 4a below. This vowel onset detection stage generates an annotation object (a Praat TextGrid) containing intervals between consecutive vowel onsets attributed in five steps. The first three steps implement a classical procedure for obtaining the amplitude envelope of the signal:

1. The speech signal is filtered by a second-order pass-band Butterworth (or Hanning) filter. This filtering is done by default in the regions of formants F1-F2 where vowels (and not consonants) have the highest energy concentration;
2. The filtered signal is then rectified. This allows changing the negative part of the signal into a positive one to obtain a positive-only amplitude envelope;
3. The rectified signal is low-pass filtered at the cut frequency of either 20 Hz (see step 4a) or 40 Hz (see step 4b) and then normalized by dividing it by its maximum value. Low-pass filtering is meant to smooth amplitude envelope and make the detection of abrupt changes in amplitude simpler. This normalized, band-specific amplitude envelope is called the beat wave. An example of a beat wave is shown in Figure 1 below.
4. A vowel onset (VO) is set either (a) at a point where the amplitude of the beat wave local rising is higher than a percentual threshold chosen by the user, or (b) at a local maximum of the normalized first derivative of the beat wave, provided this maximum is higher than a certain threshold, also chosen by the user (default = 12%). Figure 1 provides the result of a VO detection using criterion b. Each inter-VO interval is the interval corresponding to a VV unit, which is a phonetic syllable. Observe how vowel onsets are placed at rapid changes of the amplitude envelope;
5. A Praat TextGrid is then generated that contains vowel onsets boundaries.

After obtaining the vowel onset (VO) positions, the second stage of applying the SalienceDetector script consists of computing duration z-scores (z) for inter-VO intervals. This is done with the use of fixed values for reference mean (193 ms) and standard-deviation (47 ms) duration according to Equation (1), where *m* estimates the actual number of VV units between each interval generated by the BeatExtractor algorithm, which may miss vowel onsets (up to 20% from all vowels effectively present in the audio file).

$$Z = \frac{\frac{\sqrt{m}}{m}.dur - \sqrt{m}.Ref\ mean}{Ref\ SD} \tag{1}$$

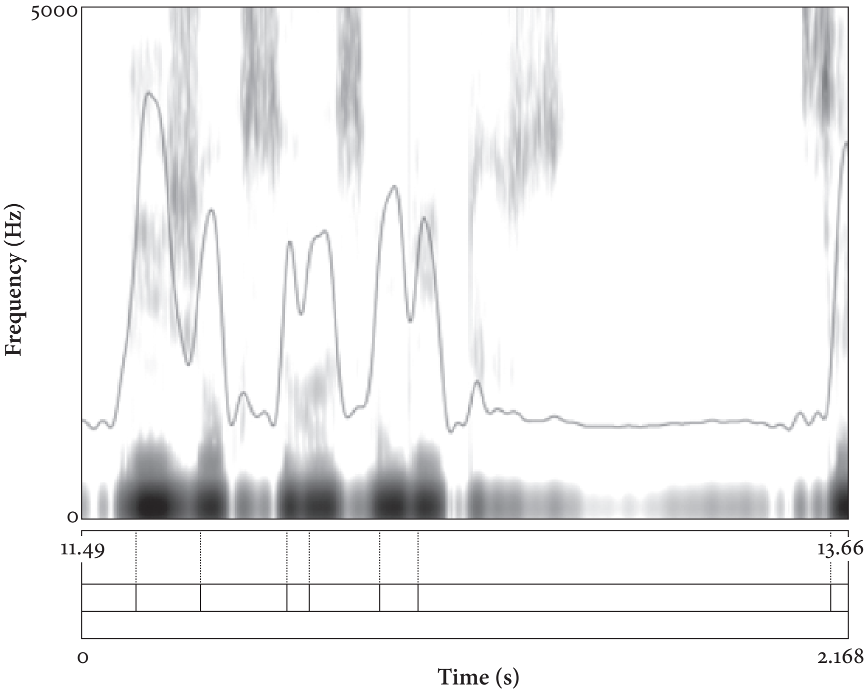


Figure 1. Beat wave (top, superposed to the spectrogram); broad-band spectrogram with automatically identified VOs shown by dotted lines (middle) and segmentation of a silent pause at 2.12 s (bottom) for French, subject CA, RE style

By means of experimental psycholinguistic experiments, Dogil and Braun (1988) have shown that vowel onset tracking by means of C-V transition detection is a fundamental property of speech signal processing in our brain. This property was also pointed out by Chistovich and Ogorodnikova (1982) by examining post stimulus temporal neuronal responses to speech. They report amplified neuronal responses to portions of energy increase typical of C-V transitions, accompanied by response suppression in regions where the energy decrease (typically around V-C transitions). That is why VV unit segmentation is a first step to capture prosodic-relevant duration variation along the utterances (Barbosa, 2006). Furthermore, a segmentation based on vowel onsets has the advantage of being detectable under moderately noisy conditions (Barbosa, 2010).

The second stage of the SalienceDetection script consists in detecting local peaks of prosodic-relevant VV durations by serially applying a smoothing technique carried by a 5-point moving average filter given by Equation (2) to the sequence of z-scores of Equation (1).

$$Z^{i}_{smoothed} = \frac{5.z^{i} + 3.z^{i-1} + 3.z^{i+1} + 1.z^{i-2} + 1.z^{i+2}}{13} \qquad (2)$$

This technique minimizes the effects of intrinsic duration and number of segments in the VV unit, as well as attenuates the effect of minor duration variation related to the implementation of lexical stress. Local peaks of smoothed z-scores are then detected by tracking the position of the VV unit for which the discrete first derivative of the corresponding smoothed z-score changes from a positive to a negative value.

The effect of the application of the two stages presented above can be seen in Figure 2, which presents only five peaks, with three of them corresponding to perceived prominence (*mosteiro* 'monastery') and boundary at the end of the words *ano* 'year' and *viver* 'to live' in the sentence *Manuel tinha entrado para o* ***mosteiro*** *há quase um* ***ano****, mas ainda não se adaptara àquela maneira de* ***viver***. 'Manuel had entered the monastery almost a year ago, but he had not yet adapted to that way of living'.

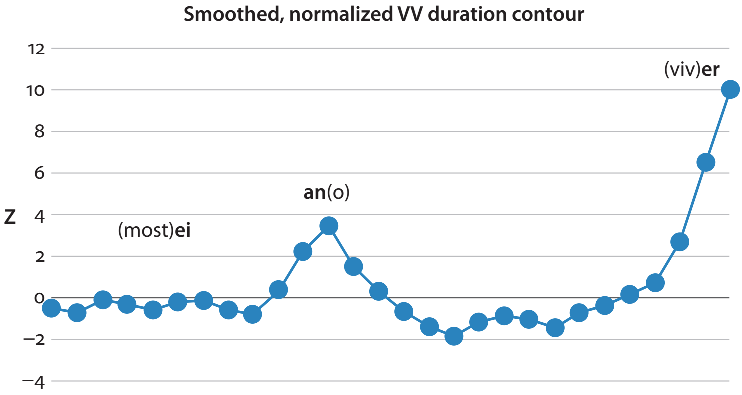


Figure 2. Smoothed, normalized VV duration contour of sentence "Manuel tinha entrado para o *mosteiro* há quase um *ano*, mas ainda não se adaptara àquela maneira de *viver*." by a female speaker

At the output, the script generates two text files, a TextGrid object and an optional plot of the syllable-sized smoothed duration along the time-course of the Sound file under analysis. The first text file is a five-column table displaying the following values for each VV unit: (a) a label which counts each unit, from s1 to s*n*, where *n* is the number of detected VVs, (b) its raw duration in milliseconds, (c) its duration z-score, the result of Equation (1) above, (d) its smoothed z-score, the result of Equation (2) above, and (e) a binary value indicating whether its position corresponds to a local peak of smoothed z-score (value = 1) or not (value = 0).

The second text file is a two-column table containing (a) the raw duration in milliseconds of the acoustically-defined stress groups, which are delimited by two consecutive peaks of smoothed z-scores and (b) the corresponding number of VV units in these stress groups. The TextGrid generated by the script contains an interval tier delimiting the detected stress group boundaries. The optional feature, implemented when the option *DrawLines* is chosen in the input parameters window, plots a trace of the smoothed z-scores synchronized with the VV unit sequence: Each value of smoothed z-score is plotted in the y-axis in the position of each detected vowel onset. This has the advantage of allowing the exam of the relation of the f0 contours with the duration-related local peaks, that is, intonation stricto sensu with rhythm stricto sensu.

The correspondence between smoothed z-scores peaks and perceived salience, which refers to both prominence and prosodic boundary, is striking. In Barbosa (2010), we demonstrated an accuracy varying from 69% to 82% between perceived prominence and boundary with produced salience.

In order to increase these figures, it is necessary to consider values of smoothed z-score peaks higher than a certain threshold that could signal the function of prosodic boundary or break, leaving aside local peaks signaling prominent VV units. After choosing this threshold we examined the correspondences between automatically detected breaks and perceived breaks, irrespective of being marks of terminality or non-terminality.

3. Results

The distribution of the peaks of smoothed z-scores for each style, for each subject in each language was grosso modo bimodal as Figure 3 illustrates for two styles in two languages. This bimodality is inferred considering the set of all histograms, though. In order to choose one single threshold for the entire corpus under analysis, the rationale behind was to assume that the lower-mode distribution can be roughly characterized by a zero-centered, normal distribution for which a z-score value lesser than 2.5 contains more than 99% of all values in the distribution. Thus, values higher than 2.5 were assumed to be part of the higher-mode distribution. We hypothesize, then, that z-score peak values higher than 2.5 signal prosodic breaks.

Thus, all breaks were generated automatically by using the SalienceDetector script and then selecting only the boundaries whose associated smoothed z-score was higher than 2.5. For doing so we used a first derivative threshold of 12% as criterion for detecting vowel onsets (see Figure 1). The results are shown in Table 1 according to language, subject and style. A hit represents the coincidence of predicted and perceived boundary at exactly the end of the word preceding the break, irrespective of being a terminal or non-terminal break. A false alarm is a predicted boundary placed in a position not perceived as a break. A miss is a perceived break

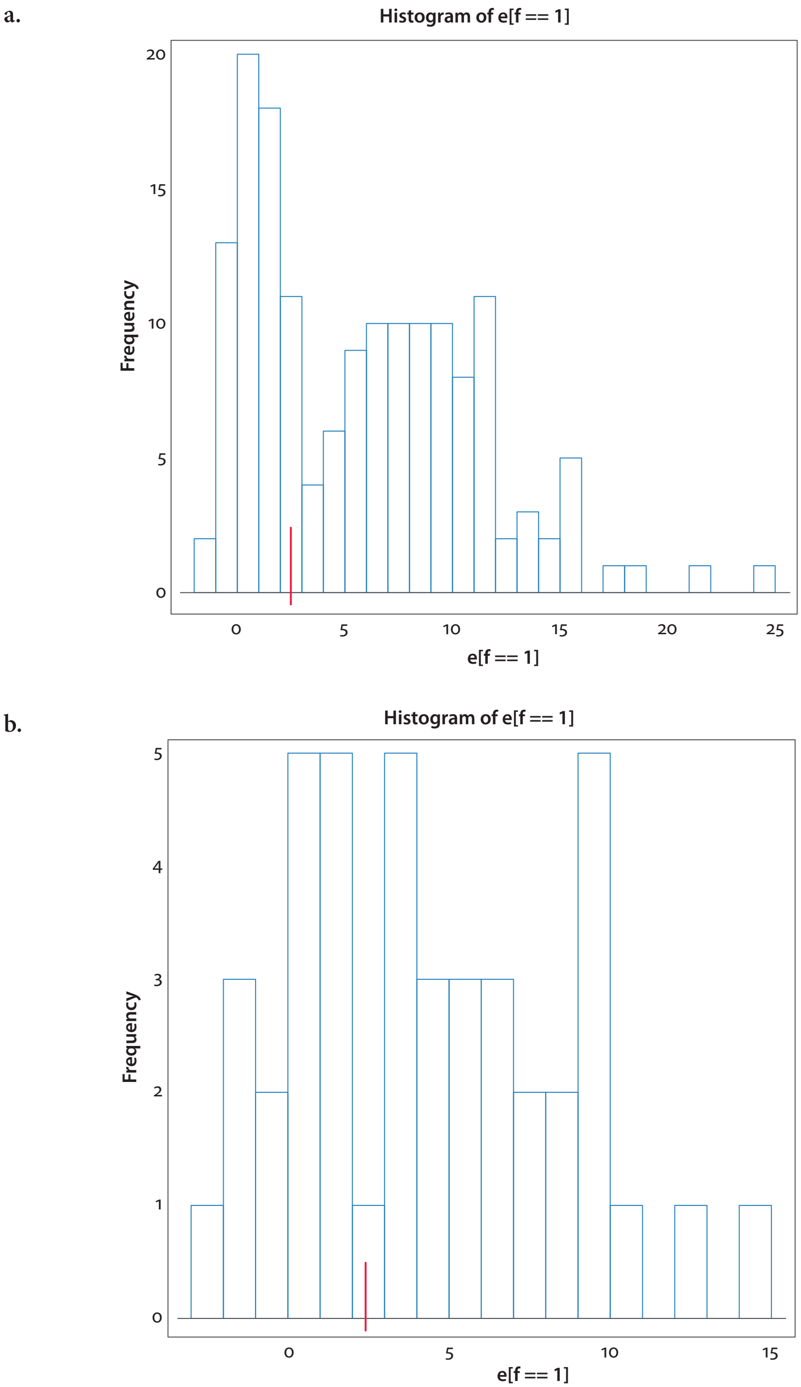


Figure 3. Histograms of smoothed z-score peaks for BP speaker NP in the reading style (a), and for French speaker MR in the narrated style (b). The short line in the two histograms represent the 2.5 value

not predicted by the algorithm and a displacement is a predicted break in a close vicinity of the perceived break. This vicinity is not larger than a phonetic syllable.

It can be seen that (1) the breaks of the Reading style are better predicted (about 70%) than those of Narration (about 60%); (2) the proportion of misses is higher for Narration; (3) EP and French have a higher proportion of hits and a lesser proportion of displacements; (4) predictions for German have a performance inferior than that for the other languages.

The reason for the worst performance of German is due to several misses in vowel onset detection, which is the first step of the algorithm. That is why a second run was carried out using a first derivative threshold of 6% as the vowel onset detection criterion. The corresponding results are shown in the same table in parentheses. This change of VO detection criterion was enough to improve break detection for German with a hit increase from 36 to 48% and a miss decrease from 35% to 27%.

Table 1. Proportions of hits, False Alarms (FA), Misses and Displacements (D) for the corpus for detection threshold = 12% (for German, also threshold = 6%). Improvements higher than 5% due to threshold changing are marked in bold

Language	Subject	Style	Hit	FA	Miss	D
BP	RA	RE	0.82	0.02	0.11	0.06
BP	RA	NR	0.71	0	0.29	0
BP	NP	RE	0.72	0.13	0.07	0.09
BP	NP	NR	0.55	0	0.3	0.15
EP	AM	RE	0.82	0.11	0.07	0
EP	AM	NR	0.68	0.18	0.12	0.02
EP	AR	RE	0.73	0.11	0.12	0.04
EP	AR	NR	0.81	0.1	0.06	0.02
FR	CA	RE	0.93	0.02	0.05	0.01
FR	CA	NR	0.6	0.13	0.27	0
FR	MA	RE	0.72	0.01	0.25	0.01
FR	MA	NR	0.76	0.06	0.15	0.03
GE	S5	RE	**0.51 (0.65)**	**0.31 (0.23)**	0.08 (0.09)	0.1 (0.03)
GE	S5	NR	**0.2 (0.31)**	0.13 (0.09)	**0.57 (0.5)**	0.11 (0.1)
GE	S6	RE	**0.3 (0.39)**	0.14 (0.19)	**0.45 (0.3)**	0.11 (0.12)
GE	S6	NR	**0.44 (0.58)**	0.06 (0.11)	**0.29 (0.17)**	**0.21(0.14)**
Total for Reading			0.69 (0.72)	0.11 (0.10)	0.15 (0.13)	0.05
Total for Narration			0.59 (0.63)	0.08	0.26 (0.23)	0.07 (0.06)
Total for BP			0.70	0.04	0.19	0.08
Total for EP			0.76	0.13	0.09	0.02
Total for FR			0.75	0.06	0.18	0.01
Total for GE			**0.36 (0.48)**	0.16 (0.16)	**0.35 (0.27)**	0.13 (0.10)

Another possible reason for the worst performance for German is related to the subjects, who have produced long stretches of speech with hypoarticulation.

3.1 Testing with English spontaneous speech

The results for German call for a test with another Germanic language. For this purpose, we analyzed the American English *Navy* dialogue used in the second part of this book. This dialogue has lesser interruptions from the interlocutor and has similarities with narrated speech. The additional reasons for using it are threefold: (a) to compare the algorithm performance in English spontaneous speech with the German narrated speech studied here; (b) to allow a comparison of automatic segmentation with human segmentation investigated in the chapters of the second part of this book; (c) to test the effect of additional criteria for detecting vowel onsets on performance improvement. The English text is transcribed below with segmentations made by Tommaso Raso. We took into account only the consensual segmentation, which in this transcription are the ones not between brackets. This gives 87 terminal or non-terminal boundaries in total.

*TOC: when I came back (/) from one of those (&he) / trips / from down to (&he) / Cartagena / I found a big stack of navy orders / (//) right //$

*TOA: hum hum //$

*TOC: so I went to this / &m what I thought was my friend (&he) / &th / this navy captain down at the naval headquarters / (//) I said / this is terribly awkward // I've just been promoted / (0) from (/) 0 third mate / to second mate / and [/] and could we / possibly / postpone these orders (/) for a little bit // my friend / stood up / behind his desk / in his full &f four stripes / and said / Lieutenant / you / are in the ues Navy now // I <said (/) oh> // (/) no one ever explained that to me before / and a week later I was on my way out to Korea //$

*TOA: <yyyy> //$

*TOA: oh //$

*TOA: where you got promoted / really rapidly / right //$

*TOC: well / no / I &he v [/] finally (&he) / I [/] you know / after / a couple months / I got promoted Lieutenant / and you know / and that sort of thing / but &he / <it was> all + I loved the Navy / (//) I really did like the <Navy> // it was just an exciting thing to do //$

*TOA: hum <hum> //$

*TOA: <hum hum> //$

*TOA: so / how many years there //$

*TOC: I stayed in the ues Navy / seventeen years and ten months // and then I was (/) forced out / because I failed a promotion to commander //$

*TOA: so / but all that service time / you put in / thirty-five / or forty <years / or> something like that / <right> //$
*TOC: <well> + <but I am &p [/] I'm on the retired list> for pay / for twenty-three years service$
*TOB: <hum hum> //$

If the same procedure described above is used for detecting boundaries with a smoothed duration z-score threshold of 2.5 but an amplitude threshold of 3% instead of using a first derivative technique with a threshold of 6%, the figures are: 40% of hits and 60% of misses with no false alarms. These figures are close to the ones for narration of the two German speakers, with the advantage of not having false alarms, although with the disadvantage of a higher proportion of misses. The reason for using the amplitude parameter criterion instead of the first derivative for detecting the boundaries is due to the fact that this English text has chunks of very low signal intensity, which makes vowel onset detection difficult with the first derivative detection technique.

If the duration z-score threshold is not used but, instead, all the local peaks of z-score are taken as boundary markers, there is an increase of false alarms (16%), a decrease of hits (33%) and of misses (51%). These figures are close to those of German subject S5. Additional factors to explain this lower performance for these two Germanic languages involve, in the material analyzed here, extreme variations of intensity, causing vowel onsets not being detected by the algorithm, but not only this. In these languages, intensity and f0 plays a similarly important role for signaling breaks. In English, Eriksson and Heldner (2015) showed that, as regards the signaling of lexical stress prominence in spontaneous speech, relative intensity has a leading role, closely followed by duration. This could explain the worst performance of English with our algorithm that uses only duration due to the fact that the same acoustic mechanisms used for signaling prominence also signal boundary. In German, Tamburini and Wagner (2007) experiments suggest that duration and intensity together are more accurate in predicting native listeners' perception of prominence.

Raso, Barbosa, Cavalcante, and Mittmann (this volume) investigated breaks in English as well. They segmented V-V units manually and used a duration normalization procedure that takes into account the label of the segments. This allows a more accurate detection of breaks, which can be seen especially in Figure 1 and Figure 2 in the aforementioned chapter.

4. Conclusions

Despite the use of a single parameter for break prediction, namely VV interval duration, the performance of our algorithm in the case of narration in French and EP is compatible with that of the literature for English as shown above, achieving 81% for narration in EP. As discussed above, the lowest performance for English and German is likely to be related to the similar role of intensity for signaling breaks in those languages as well as to idiosyncratic aspects of the material that made the detection of vowel onsets more difficult. Further investigation of these two languages is necessary before deciding for which languages the proposed algorithm is more indicated. The role of other parameters for signaling breaks in BP was recently investigated by Teixeira, Barbosa, and Raso (2018), which showed that f0 also plays an important role in predicting terminal breaks.

The main advantage of the algorithm proposed here is certainly the low computational cost. Additional advantages are: (1) it does not require any kind of previous linguistic analysis or transcription, which makes it more readily applicable than others in the literature that requires previous linguistic segmentation; (2) it is language-independent; (3) its input parameters can be adjusted to achieve a better performance depending on language and subject; (4) it allows to consider different levels of strength between breaks by using the smoothed z-score as a measure of break strength, which has application for the study of syntax-prosody interface.

References

Avanzi, M., Lacheret, A., & Victorri, B. (2008). Analor, un outil d'aide pour la modélisation de l'interface prosodie-grammaire. *Travaux Linguistiques du CerLiCO*, 21, 27–46.

Barbosa, P. A. (1994). Caractérisation et génération automatique de la structuration rythmique du français (Unpublished doctoral dissertation). Institut National Polytechnique de Grenoble, France).

Barbosa, P. A. (1996). At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration: Emphasis on segmental duration generation. *Cadernos de Estudos Linguísticos*, 31, 33–53.

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

Barbosa, P. A. (2007). From syntax to acoustic duration: A dynamical model of speech rhythm production. *Speech Communication*, 49, 725–742. https://doi.org/10.1016/j.specom.2007.04.013

Barbosa, P. A. (2010). Automatic duration-related salience detection in Brazilian Portuguese read and spontaneous speech. *Proceedings of the Speech Prosody 2010 Conference*, 10–14 May, Chicago, IL.

Boersma, P. & Weenink, D. (2017). *Praat: Doing phonetics by computer* (Version 6.0.29) [Computer software]. Retrieved from <www.praat.org>

Botinis, A., Granström, B., & Möbius, B. (2001). Developments and paradigms in intonation research. *Speech Communication*, 33, 263–296. https://doi.org/10.1016/S0167-6393(00)00060-1

Campbell, N. (1993). Automatic detection of prosodic boundaries in speech. *Speech Communication*, 13(3–4), 343–354. https://doi.org/10.1016/0167-6393(93)90033-H

Chistovich, L. A., & Ogorodnikova, E. A. (1982). Temporal processing of spectral data in vowel perception. *Speech Communication*, 1, 45–54. https://doi.org/10.1016/0167-6393(82)90007-3

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

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *J. Phon*, 26, 145–171. https://doi.org/10.1006/jpho.1998.0070

Eriksson, A., & Heldner, M. (2015). The acoustics of word stress in English as a function of stress level and speaking style. *Proc. of the 16th Annual Conference of the International Speech Communication Association (INTERSPEECH 2015)*, Dresden, Germany, 41–45.

Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1, 517–520.

Gotoy, Y., & Renals, S. (2000). Sentence boundary detection in broadcast speech transcripts. *Proc. of the International Speech Communication Association (ISCA) Workshop: Automatic Speech Recognition: Challenges for the New Millennium* (ASR-2000), Paris.

Kim, J. (2004). Automatic detection of sentence boundaries, disfluencies, and conversational fillers in spontaneous speech (Unpublished doctoral dissertation). University of Washington. Retrieved from <https://ssli.ee.washington.edu/papers/grad/theses/jkim-ms-thesis.pdf>

Lacheret-Dujour, A., Simon, A., Goldman, J., & Avanzi, M. (2013). Prominence perception and accent detection in French: From phonetic processing to grammatical analysis. *Language Sciences*, 39, 95–106. https://doi.org/10.1016/j.langsci.2013.02.007

Mettouchi, A., Lacheret-Dujour, A., Silber-Varod, V., & Izre'el, S. (2007). Only prosody? Perception of speech segmentation in Kabyle and Hebrew. *Nouveaux Cahiers de Linguistique Française*, 28, 207–218.

Mittman, M. M., & Barbosa, P. A. (2016). An automatic speech segmentation tool based on multiple acoustic parameters. *CHIMERA. Romance Corpora and Linguistic Studies*, 3(2), 133–147.

Ni, C. J., Zhang, A. Y., Liu, W. J., & Xu, B. (2012). Automatic prosodic break detection and feature analysis. *J. Comput. Sci. Technol.*, 27, 1184–1196. https://doi.org/10.1007/s11390-012-1295-z

Raso, T., Barbosa, P. A., Cavalcante, F. A., & Mittmann, M. M. (this volume). Segmentation and analysis of the two English excerpts: The Brazilian team proposal. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Shriberg, E., Stolcke, A., Hakkani-Tür, D., & Tür, G. (2000). Prosody-based automatic segmentation of speech into sentences and topics. *Speech Communication*, 32(1), 127–154. https://doi.org/10.1016/S0167-6393(00)00028-5

Tamburini, F., & Wagner, P. (2007). On automatic prominence detection for German. *Proc. of the 8th Annual Conference of the International Speech Communication Association (INTERSPEECH 2007)*, (pp. 1809–1812). Antwerp, Belgium.

Teixeira, B., Barbosa, P., & Raso, T. (2018). Automatic detection of prosodic boundaries in Brazilian Portuguese spontaneous speech. In A. Villavicencio, M. Viviane, A. Abad, H. Caseli, P. Gamallo, C. Ramisch, H. R. Gonçalo Oliveira & G. H. Paetzold (Eds.), *Computational processing of the Portuguese language. PROPOR 2018* (pp. 429–437). Canela, Brazil. Cham: Springer. https://doi.org/10.1007/978-3-319-99722-3_43

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *J. Acoust. Soc. Am.*, 91, 1707–1717. https://doi.org/10.1121/1.402450

Part II

Same texts, different approaches to segmentation

An introduction to the second part of the volume

Shlomo Izre'el[i], Heliana Mello[ii], Alessandro Panunzi[iii] and Tommaso Raso[ii]

[i]Tel Aviv University / [ii]Federal University of Minas Gerais / [iii]University of Florence – LABLITA

1. Introductory notes to the second part of the volume

The second part of the book has been conceived around the idea of giving practical shape to the comparisons between different theories and approaches to spoken language segmentation and the definition of a basic unit for its analysis. To this end, we collected different segmentations and respective annotations (following different theoretical frameworks as reflected in the respective chapters of Part I) for the same two American English texts, which are presented at the end of these introductory notes.

The work proceeded in two sequential phases, during which the list of annotators (or annotation teams) slightly changed. In the first phase, seven annotators provided a coarse-grained annotated version of both texts. The segmentations have been stored in the SLAC (*Spoken Language Annotation Comparison*) database (Panunzi et al., this volume), which has been specifically developed for the purpose of comparing them from a quantitative point of view.[1] All the independent segmentations have been compiled in a synoptic grid, allowing for an easy but analytical overview of similarities and differences between the annotators' perspectives. The SLAC database has been designed and developed by Alessandro Panunzi, Lorenzo Gregori and Bruno Rocha, the authors of the comparative paper at the end of this part. Both the architecture of the database and its use are described in detail within their paper (specifically in Section 3). A user guide containing the

[1] Freely accessible online at <http://lablita.it/app/slac/>

https://doi.org/10.1075/scl.94.int2

basic instructions is also present in the online SLAC interface. The segmentations have been analyzed in detail and quantitatively compared during the 2015 meeting entitled “Units of Reference for Spontaneous Speech Analysis and their Correlation across Languages”, organized by LABLITA and LEEL at the Federal University of Minas Gerais in Belo Horizonte, Brazil. This initial work was the impetus to make a further, more challenging step in the paper now presented in this volume.

For the second phase of the comparison, the participants decided to refine their annotations and to collect them in a joint publication. The papers included in the second part of this book are the result of this second phase. The authors described the work done, explaining with further detail the criteria and annotation schema used, and additionally wrote theoretical notes focusing on different aspects. These include, for instance, how they conceive the basic unit of the spoken language, how the basic unit relates to other levels of analysis (pragmatic, syntactical, discursive), how the prosodic units can be internally structured, or how the perceptive boundaries can be described by means of formal features. Prosody comprises the basis for every annotation schema, and each team attributes great importance to it as regards the delimitation of pragmatic-discursive units.

Most papers further include a fine-grained annotation of two shorter excerpts taken from each text, which makes it useful to compare in more detail the annotation frameworks and the specific procedures followed during the development of the task. At the end of each paper, the appendices schematically report the categories and tags used in the analysis, and (where a fine-grained analysis has been performed) tables summarize the in-depth analysis of the smaller excerpts, with all the different levels considered and the parameters extracted.

It is worth noting one more time that the list of annotators who performed the two steps of the comparative work is not exactly the same. The authors of full segmentations stored in the SLAC database are (abbreviations in brackets): Chafe (CHA); Cresti-Raso (CNR); Izre’el (IZR); Kibrik-Korotaev-Podlesskaya (KKP); Martin (MRT); Maruyama (MAY); and Mithun (MIT).

In the comparison paper by Panunzi, Gregori and Rocha these abbreviations are used to refer to the segmentations stored in the SLAC database, while author names (without further specifications) are used for referring to the theoretical explanations and to the fine-grained analyses included in the papers of the second part of this book. The list of authors is: Raso, Barbosa, Cavalcante, and Mittmann (Raso et al.); Martin; Izre’el; Mithun; Maruyama; Kibrik, Korotaev, and Podlesskaya (Kibrik et al.); Cresti and Moneglia.

The comparison paper starts by illustrating the different theoretical perspectives with respect to three main aspects: (1) the role of prosody in segmenting the speech flow; (2) the relation between prosody and syntax; (3) the definition and the nature of the reference units for analyzing the spoken language. After this, the

SLAC database is presented in detail, explaining its content and how to use it. The database web interface contains: (1) the unannotated transcripts of both texts, divided in dialogic turns; (2) for each turn, list of single segmentations (inline view) and a synoptic grid in which it is possible to compare all of them. The sound files can be played directly from the web interface, allowing a real-time comparison of the different segmentations *vis-à-vis* the source data. The paper also introduces the Unified Tagset, that is, an integrated representation of all the different annotation schemas used by each team, which have been reported as regards the main prosodic distinction between terminal and non-terminal prosodic boundaries. Finally, the paper reports the quantitative analysis of the agreement between segmentations, both as an overall comparison (i.e., all annotations taken together) and as a series of pairwise contrasts. The results of these comparisons highlight the fact that the identification of prosodic breaks (notably terminal ones) is mostly based on general segmentation abilities based on perceptual factors, which turn out to be largely independent from the theoretical framework adopted. This perceptual segmentation should thus be viewed as a basic, non-theory-bound level of annotation that should be present in all representations of speech data.

Given the above, we hope that the work done can constitute a starting point for the development of a shared tagset for the annotation of perceptive boundaries in spontaneous speech analysis.

2. The analyzed texts

The texts on which each annotation team worked have been extracted from the Santa Barbara Corpus, and are titled *Hearts* and *Navy*.[2] *Hearts* is mostly dialogical and interactive, and contains an exchange in which one speaker explains to the other how to play the card game called Hearts. On the other hand, *Navy* is more monological and narrative: There is a main speaker telling about his military career in the U.S. Navy, and the text includes only minimal contributions by two other speakers, who mainly play the role of listeners.

This choice of texts reflected two main requirements: The first one is to have a good acoustic quality for speech analysis, and the second is to represent different types of interaction. This latter fact has direct consequences on the segmentation task, since dialogic texts tend to produce shorter turns and shorter prosodic units, with a richer internal variation, while monologic texts tend to produce longer units with less variation. On the other hand, syntax seems to have a stronger role in

[2] Santa Barbara annotated minicorpus available at: <http://www.c-oral-brasil.org/> Corpora > American English minicorpus>.

structuring monologic interactions than dialogic ones, while dialogic exchanges tend to have a substantial number of short interpretable units with limited syntactic structure (e.g., verbless utterances).

For the annotation comparison task, each team has been provided with the sound file corresponding to the texts, and the bare orthographic transcripts reported below. The only symbols used are: "&" at the beginning of each word fragment; "yyyy" for non-vocal audio signals. The *Hearts* sound file is 1.41 min long, and its bare transcription comprises 27 turns and 363 words. The *Navy* sound file is 1.17 min long, and its bare transcription comprises 18 turns and 251 words. The text of the shorter excerpts used for the in-depth analysis is highlighted in bold.

Hearts

1. *DAN: what's hearts
2. *JEN: hearts it's the card game
3. *DAN: oh yeah put it up there
4. *JEN: wanna play hearts
5. *DAN: let's check that one out **neat wait play novice I've never played hearts before in my life**
6. *JEN: **you've never played hearts**
7. *DAN: **no I don't know how to play it**
8. *JEN: **oh okay I'll teach you**
9. *DAN: **passing disabled**
10. *JEN: **queen of &sp**
11. *DAN: that's you
12. *JEN: &he first lead rotates first yeah always pass left alright so this is us okay every heart is one point the &q queen of spades is thirteen points the object is not to have any points
13. *DAN: is &tr
14. *JEN: and you play following suit and you can take if you take tricks &th the highest card of the suit takes the trick if you don't have the card of the suit you throw whatever you want
15. *DAN: okay so &h hearts and the queen of spades are the only thing that that have points
16. *JEN: are bad that are that are points right
17. *DAN: so we got like three points right here right
18. *JEN: we have three points in our hand exactly
19. *DAN: and we &w wanna try to get rid of that
20. *JEN: right but we're passing now the first thing you do is you pass three cards to your left now these are low hearts so I can I'm not gonna pass those I'm gonna pass the &f four of clubs and these are two high
21. *DAN: all the &y &right why is that

22. *JEN: why just because and cause you should always pass a club so that the person so the first hand everyone has a club so that they can't discard a heart cause you always assume that everyone's &t no one is void of a suit the first time around so you don't have to worry about throwing a high card
23. *DAN: yeah yeah
24. *JEN: and then I'm gonna throw two high cards so I don't take those tricks because
25. *DAN: but what difference does it make if you take a trick
26. *JEN: well because &I cause if I &t if I take a &tr the &k diamond trick and somebody didn't have diamonds and they threw a heart into that pile I was gonna take that with that ace
27. *DAN: they're not worth anything

Navy

1. *TOC: when I came back from one of those &he trips from down to Cartagena I found a big stack of navy orders right
2. *TOA: hum hum
3. *TOC: **so I went to this &m what I thought was my friend &he &th this navy captain down at &he naval headquarters I said this is terribly awkward I've just been promoted from third mate to second mate and and could we possibly postpone these orders for a little bit my friend stood up behind his desk in his full &f four stripes and said Lieutenant you are in the ues Navy now** I said oh
4. *TOA: yyyy
5. *TOC: no one ever explained that to me before and a week later I was on my way out to Korea
6. *TOA: oh
7. *TOC: yeah
8. *TOB: where you got promoted really rapidly right
9. *TOC: well no I &he &v finally &he I you know after a couple months I got promoted Lieutenant and you know and that sort of thing but &he it was all
10. *TOA: hum hum
11. *TOC: I loved the Navy I really did like the Navy
12. *TOA: hum hum
13. *TOC: it was just an exciting thing to do
14. *TOB: so how many years there
15. *TOC: I stayed in the ues Navy seventeen years and ten months and then I was forced out because I failed a promotion to commander
16. *TOB: so but all that service time you put in thirty-five or forty years or something like that right
17. *TOC: well but I am &p I'm on the retired list for pay for twenty-three years service
18. *TOA: hum hum

CHAPTER 1

Segmentation and analysis of the two English excerpts

The Brazilian team proposal

Tommaso Raso, Plínio A. Barbosa, Frederico A. Cavalcante and Maryualê M. Mittmann
Federal University of Minas Gerais, CNPq, FAPEMIG / State University of Campinas, Institute for Language Studies, CNPq / Federal University of Minas Gerais, CAPES / Unifacvest

This paper has a tripartite focus: (1) to establish the best segmentation for two American English texts according to inter-rater agreement measurements. By doing this, we differentiate the behavior of experts and non-experts annotators. The experts' annotation constitute the basis for the analysis; (2) to capture, measure, and analyze the phonetic features that correlate with boundaries, as they are marked by the expert annotators; (3) to informationally annotate prosodic units according to the Language into Act Theory, and analyze their corresponding information structure: in order to do this, we make and justify decisions in marking the reference units and assigning informational value to prosodic units; additionally we further discuss some cases of major disagreements.

Keywords: speech segmentation, boundaries, phonetic features, inter-rater agreement, information structure annotation

1. Introduction

We assume, in line with many other scholars, that speech is prosodically segmented into intonation units (IU). By IUs we mean prosodic organizations that encapsulate a certain amount of segmental material. IUs are separated by boundaries that are generally clearly perceivable. Their perception is cued by combinations of acoustic parameters, but these combinations are yet to be determined (Barbosa & Raso, 2018), although some parameters are commonly considered in the literature as playing a strong role in cueing boundary perception. A full list should include, at least, pause, f0 parameters, duration parameters, intensity parameters, rhythmic parameters, and voice quality change.

https://doi.org/10.1075/scl.94.10ras

Prosodic boundaries, however, are not always of the same type. We must distinguish at least between conclusive and continuative boundaries. Different kinds of continuative boundaries can probably be found; for example, some presenting a specific continuative tone, and some lacking such a clear tone, but nonetheless perceived as non-conclusive (Teixeira, Barbosa, & Raso, 2018).[1]

Despite the rich literature on the prosodic segmentation of speech (see Barth-Weingarten, 2016, for a survey), it is clear that we still need to improve our methodology in order to completely capture this phenomenon and its physical correlates. In this paper, we try to give a contribution in a very interesting exercise together with other scholars from different traditions and native languages, all of us concentrating on the same excerpts of two English texts, one dialogic and one monologic (see Introduction to Part II). We will proceed as follows:

1. We will present and discuss a perceptually-based segmentation of both texts by 16 annotators;
2. We will show some measurements that could acoustically justify the results of the perceptually-based task;
3. We will propose a functional analysis of the different units that emerged from the segmentation.

Our attempt was primarily to segment the texts without any theoretical influence (as far as this is possible), and only then to analyze the functional effects of the segmentation. This means that we take IUs as having a mainly functional motivation in the organization of speech.

2. Inter-rater agreement in the segmentation

To decide how to segment the excerpts of the texts *Hearts* and *Navy* extracted from the *Santa Barbara Corpus* (Du Bois, Chafe, Meyer, & Thompson, 2000–2005), we asked 16 annotators to perform the task. Among them, there were six expert annotators with good or very good knowledge of English; the remaining ones were either less or not skilled in speech segmentation, or they had less competence in

[1] The possibility that we should distinguish among different types of non-conclusive boundaries is supported by the first results of a project in which two authors of this paper participate. The project aims at creating an automatic tool for spontaneous speech segmentation trained through human segmentation (see Mittmann & Barbosa, 2016; Teixeira, 2018; Teixeira et al., 2018). The partial results presently point that it is much easier to produce one unique model for detecting conclusive boundaries (so far, the model reaches an agreement higher than 80% with human annotators using less than 10 measurements) than for continuative ones (for which the agreement is much lower – around 45% – if we use just one model, but reaches almost 100% with the use of three different models, with 8 to 10 measurements).

English, or both. The perceived breaks we retagged as follows: terminal (conclusive) prosodic boundaries with two slashes (//); non-terminal (continuative) boundaries with one slash (/); non-terminal ones due to fragmentation phenomena with one slash within brackets and followed by a digit indicating the number of retracted words ([/n.]); in addition, the ampersand sign (&) precedes time-takings and interrupted words, and angle brackets (<>) indicate overlapping speech.

We evaluated the inter-rater agreement in the segmentation task through the Kappa statistic (Fleiss, 1971), whose results are shown in Table 1. We also calculated the Kappa disregarding the fragmentation phenomena, which did not yield any significant change in scores. Regarding the dialogic text (*Hearts*), we reached a general Kappa of 0.82, which is considered excellent, for the 16 annotators, and a general Kappa of 0.90 for the six expert annotators. The agreement on terminal breaks was of 0.80 for the 16 subjects and of 0.88 for the expert ones. The agreement on the non-terminal breaks was of 0.40 for the 16 subjects and of 0.62 for the six experts.

These results suggest that expertise increases the judgment of non-terminal break significantly, since 0.40 is not a good agreement, while 0.62 is. We also did a test to evaluate the agreement regardless of the nature of the boundaries (i.e., terminal or non-terminal). The results produced a Kappa of 0.94 for all the 16 subjects and one of 0.98 for the six experts. These scores confirm the impression that it is easy for any annotator, even untrained ones, to decide between presence versus absence of boundary, while it is more difficult to decide about the boundary's nature. Expertise seems to play an important role for this kind of decision.

Table 1. Inter-annotator agreement

Text	Annotators	General kappa	Terminal	Non-terminal	Annotators	Kappa break *vs.* no break
Hearts	16	0.82	0.80	0.40	16	0.94
	Top 6	0.90	0.88	0.62	Top 6	0.98
Navy	16	0.71	0.72	0.59	16	0.76
	Top 6	0.76	0.91	0.71	Top 6	0.76

The general Kappa for the monologic text was 0.71 for all the annotators and 0.76 for the six experts; in both cases the results were very good. The agreement for terminal boundaries was 0.72 for all the annotators and 0.91 for the six experts; the agreement for non-terminal boundaries was 0.59 for all the annotators and 0.71 for the six experts. Also, for this text we did a test challenging only the difference between presence versus absence of boundary, and we reached the same result, that is, 0.76, for both groups. Again, we can conclude that the decision regarding the presence versus absence of a boundary does not seem to require a special expertise, whereas expertise does seem to be important to decide the nature of a boundary.

3. Phonetic measurements

Our final segmentation was done according to those of the six expert annotators (Appendices A and B). We will make some comments in case of strong disagreement across annotations. To better understand the possible acoustic cues that guides an annotator's perception, we took the following measurements, reported in Appendices A and B:

1. Smoothed z-scores of normalized, pre-boundary VV durations;
2. Presence and duration of pauses;
3. Difference of global intensity (dB) means taken from vowels of boundary syllables. According to Senn, Kompis, Vischer, and Haeusler (2005), the just noticeable difference (JND) is 1 dB;
4. Spectral emphasis of the same syllables, to avoid interferences due to change in position of microphone during recording;
5. Difference between the first f0 value (semitone) at the right of the boundary and the last one at the left of the boundary;
6. Difference between f0 means (semitone) of vowels of boundary syllables. According to t'Hart (1981), the JND is 3 semitones, but it can change depending on the context and on the interaction with other features;
7. Change of f0 movement across boundary;
8. Difference in articulation rate between intonation units across the boundary;

Concerning the difference in articulation rate (8), for the units with more than four syllables, we also calculated the difference in articulation rate considering just the last four syllables at the left and/or the first four syllables at the right of the boundary. In no case did these differences turn out statistically significant, but this may be due to insufficient number of measurements; in fact, according to Quené (2007) and other studies, the JND for articulation rate is 5%. As Appendices A and B show, all the boundaries in the two texts feature differences higher than 5% in articulation rate between the unit at the left and the unit at the right (or between the last four syllables at the left and the first four at the right). This allows for us to say that articulation rate can often be considered an important feature in determining perception of boundaries.

In addition, stress groups were automatically located using the *SG_Detector* script (Barbosa, 2013), which provides the normalization (smoothed z-scores) of the durations of VV units. The reference durations (means and standard deviations) for the English segments, based on British English, were taken from Campbell (1992).

Figure 1 shows the relation among intonation unit boundaries (breaks), duration peaks (stress group boundaries), and pauses for the text *Navy*. Intonation unit boundaries are signaled in red; the highest peaks indicate terminal boundaries, while the intermediate and lowest ones indicate non-terminal boundaries and boundaries produced by fragmentation phenomena respectively. The normalized duration curve is indicated in yellow; the different heights of the peaks reflects the normalized durations of the VV units. When the boundaries include a pause, a blue vertical line is added; the size of the blue bar is proportional to the pause duration. The numbering on the x-axis refers to the VV units into which the text was segmented to calculate stress groups and normalized duration. Figure 2 shows the same measurements for the text *Hearts*.

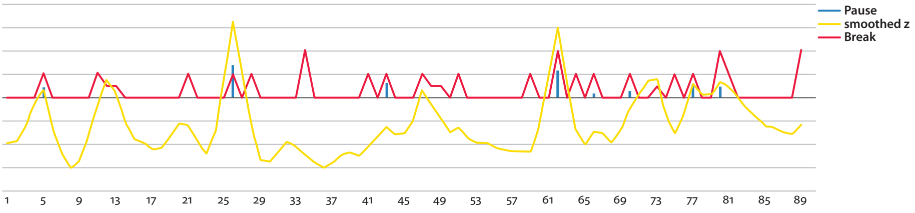


Figure 1. IU boundaries, pauses, and normalized durations curve (smoothed z) in *Navy*

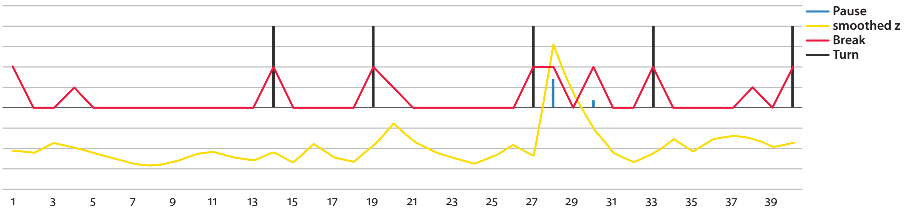


Figure 2. Turn and IU boundaries, pauses, and normalized durations curve (smoothed z) in *Hearts*

As Figure 1 shows, there is a high correspondence between durational peaks and intonation unit boundaries. Sometimes this relation is stronger, like in VV units 5, 12, 26 (which includes a pause), 47, 62 (which includes a pause too), 73, 77, 80; and sometimes it is less strong, like in 20, 32, 39, 51, 66, 89. Sometimes the durational peak is one or two syllables before the intonation boundary, like in 20–21, 32–34, 39–41. For a small number of unit boundaries, we do not find any durational peaks (28, 41, 59, 64, 70, where there is a pause). The situation is not much different in the dialogic text, as Figure 2 shows.

These correspondences suggest that durational peaks can play an important role in cueing the perception of IU boundaries, but they do not seem to correspond entirely to IU boundaries. There can be durational peaks that do not correspond exactly to IU boundaries, as well as IU boundaries without durational peaks. For the first case, we still can think that the durational peak "warns" that there is a boundary, but the latter will be phonologically positioned at the end of the phonological word.

Naturally, when a pause is present, its position always corresponds to an intonation boundary. Nevertheless, neither presence nor duration of pause seem to be strongly related with the nature (terminal or non-terminal) of the boundary. There is, for instance, a long pause in 26 (Figure 1), where a non-terminal boundary was marked, but a short pause in 80 and no pause in 34, where terminal boundaries were marked (see Raso, Mittmann, & Oliveira Mendes, 2015, for an in-depth analysis of the relation between pause and prosodic boundary).

Other measurements seem to cooperate in cueing boundary perception, as the tables in Appendices A and B show. We consistently found a generally significant difference in intensity between the last vowel of the left IU and the first vowel of right IU, as well as a JND between articulation rates. Differences in f0 between the last point at the left of the boundary and the first point at the right of the boundary are less common (less than 1/3 of the cases), which suggests that f0 reset is an important feature for cueing the perception of boundary, but is not at all sufficient to explain boundary perception. As for change of f0 movement, only in three cases did we find a clear change across the boundary.

Another interesting parameter to observe is the change of voice quality close to a boundary, especially terminal ones. In *Hearts*, we observe a change of voice quality in *okay*, which ends with breathy voice, and in *I'll teach you*, which shows creakiness. In *Navy*, there is creaky voice at the end of *possibly postpone these orders for a little bit* and breathy voice at the end of the last unit.

Regarding the segmentation, there are only two points of relevant disagreement among the annotators, both in the monologic text. The first one is in the units 10 to 11 (Appendix B). In this case, three of the top six annotators did not place a boundary after *from*, while two did place a boundary after *from* and one before *from*. We decided to maintain the boundary after *from* because all acoustic measurements show there is great motivation for boundary perception and, from the functional point of view (see Section 4.2), it is very likely that we have two different units. The second disagreement is relative to the unit 16 to 17. Here, most of the 16 annotators (11), placed a boundary after *orders*, but four of the six expert annotators identified no boundary at all. Likewise, we decided not to consider the boundary in this case because the measurements suggest there is no relevant acoustic motivation for its perception and because, from the functional point of view, there is no strong evidence for a second separated unit (see Section 3.2).

4. Reference unit and IUs: A functional analysis

4.1 The reference unit

For the functional analysis, we follow the principles of the *Language into Act Theory* (L-AcT; Cresti, 2000; Moneglia, 2005; Moneglia & Raso, 2014; see also Bossaglia, Mello, & Raso, this volume; Cresti, this volume; Cresti & Moneglia, this volume). We consider as the communicative reference unit for speech what we can call *terminated sequence* (TS), defined as the smallest stretch of speech that is both pragmatically and prosodically interpretable in isolation. The pragmatic interpretability is conveyed by the presence of at least one illocutionary unit; the prosodic interpretability is conveyed by a boundary that yields the perception of conclusion.

The Kappa scores for the segmentations showed that the very strong agreement concerning the presence versus absence of a boundary becomes less evident when the annotators must decide about the nature of the boundary. This could be interpreted as a weakening of the concept of reference unit based on the perception of terminality. We will come back to this point.

TSs can coincide with one IU or with a sequence of more than one IU. They coincide with one IU when the IU carries the illocutionary force (and therefore the pragmatic interpretability) and its boundary conveys conclusion. We can call these kind of TS "simple utterances". We found five such cases in the dialogic text, *Hearts*: 1. *Wait //*; 4. *<you've never> played hearts //*; 7. *oh //*; 8. *okay //*; 9. *I'll teach you //*.

All the TSs of the monological text, *Navy*, and the other TS of *Hearts* should be considered complex sequences, since they are formed by more than one IU, according to most of our annotators. We can distinguish between two different kinds of complex TS: *complex utterances* and *stanzas*. Complex utterances show only one pattern, which can vary in complexity. In this case, we find one illocution (or two or more *patterned* illocutions, as we will show) with other non-illocutionary units that form a pattern around the illocution. Stanzas are formed by more than one pattern: Each pattern is made up of one illocution (mandatory) and non-illocutionary optional units that prosodically depend on a specific illocution; the different patterns are juxtaposed and separated by a continuation prosodic tone at the boundary (or by a prosodic break that does not convey the perception of terminality) (Cresti, 2009).

Most of the disagreement on the nature of the boundary (conclusive/terminal vs. continuative/non-terminal) coincide with prosodic boundaries that do not feature a clear continuation tone, which is conveyed by at least a rising movement and a final lengthening.[2] These boundaries are positioned after the accomplishment of an illocutionary unit, but they exhibit prosodic cues that, although lacking a clear

[2] See Silber-Varod (2011) for an interesting description of five different continuative boundaries.

continuation tone, nevertheless do not convey the perception that the unit is fully concluded. Once the illocution is accomplished, this kind of boundary can easily lead to disagreement in the assessment of the boundary. In fact, the annotator, especially a non-trained one, can either consider the sequence as a concluded one, since no continuation tone is perceived, or place a non-terminal boundary, since, despite the lack of a continuation tone, conclusion is not fully perceived. These cases are those with less agreement among the annotators. Therefore, we can say that, to perceive a terminated sequence, it is mandatory to have the accomplishment of the illocution; but when we have neither a clear continuation tone nor a strong reason to perceive terminality (end of a turn or a clear final profile), the perception of complete conclusion can easily be subjective. In our excerpts, a good example of this kind of potentially ambiguous break is the boundary after unit 7 in *Navy* (*naval headquarters*). Here, no explicit continuation tone is present, but at the same time, the falling profile ends higher than the lowest f0 level of the unit. Moreover, the prominence on the first syllable of the last word (HEADquarters) can interfere with the perception of the boundary. It is not surprising that this is the break that exhibits the highest disagreement (either terminal or non-terminal). No other position presents such a strong uncertainty between terminal and non-terminal boundary. Interestingly, the disagreement is not strong among the six experts (5 non-terminal vs. 1 terminal), but is high among the other 10 (6 vs. 4), with a preference for a solution different from that chosen by the experts. A different, though interesting, case is unit 9 (*this is terribly awkward*). Here, f0 ends at a very low level, but no syllabic lengthening is featured; additionally, this may lead to uncertainty with respect to the perception of the terminality or non-terminality of the boundary. Nevertheless, in this case, only three annotators did not chose to mark terminality. These observations are intended to explain disagreements related to the nature of specific boundaries, and can suggest that the distinction between terminal and non-terminal boundaries be insufficient. We probably need to distinguish at least among different kinds of non-terminal boundaries (for a discussion, see Barbosa & Raso, 2018; for a first attempt to model terminal and non-terminal boundaries, see Teixeira & Mittmann, 2018, and Teixeira et al., 2018).

We find the following complex patterns in *Hearts*: 2–3. *play novice / I've never played hearts before <in my life> //*; 5–6. *No / I don't know how to play* it //; 10–11. *passing disabled / <that's you> //*. As we can observe, their complexity is limited to just two IUs. None of them comprises a *stanza*.

The situation in the monological text *Navy* is much different. Not only do we have mostly complex sequences, but, what is more, these sequences are usually more complex than those in *Hearts*. We can divide them into complex utterances and *stanzas* (we will come back to the difference later). The only complex utterance seems to be the last one: 25–26. *you / are in the ues Navy now //*. All the other sequences are *stanzas*: 1–9. *so I went to this / &m / what I thought was my friend / &he / &he / this navy captain down at the / naval headquarters / I said / this is terribly awkward //*; 10–17. *I've just been promoted from / third mate / to second mate / and / and / could we / possibly postpone these orders for a little bit //*; 18–24. *my friend / stood up / behind his desk / in his full &f / four stripes / and said / Lieutenant //*

These differences between the two texts reflect well the common structural distinction between interactive dialogic texts and monologues.

4.2 Intonation units and information pattern

Functionally, IUs convey informational values. Therefore, what we propose is that the principle that guides the grouping of words in IUs is the informational value that the speaker assigns to them. Any specific informational value rests on a specific prosodic form that cues the perception of the specific information value. We will describe the features of information units in the two texts.

First, we can focus on the illocutionary units, which are the only informational value that cannot be lacking for a TS to exist, since on the illocution rests the communicative power of a TS, that is, its property of being a speech act. All simple utterances coincide with illocutionary units. The main prosodic feature of these units is a functional focus (Cresti, 2011), that is, a prosodic prominence that is responsible for conveying a specific illocutionary value. This is what can be called the illocutionary nucleus. Usually, it occupies only one or two syllables of the unit; the other syllables, if any, are responsible for carrying the syllabic content of the locution (Cresti, 2018; Raso & Rocha, 2017; Rocha & Raso, 2016).

In *Hearts*, all the IUs are illocutionary, even those in the complex utterances. In fact, in these cases complex utterances are formed by couples of patterned illocutions (Multiple Comments – CMM) that yield a holistic interpretation as a rhetorical pattern. In our case, 2–3 and 5–6 produce a pattern of reinforcement, with the repetition of the same illocutionary value in both units; the illocutionary nuclei are realized by the first syllable of *novice, hearts, no* and *play it*. In 10–11 the pattern conveys a relation of cause and effects; this relation is conveyed prosodically, and the prominences are, respectively, in the last two syllables of *disabled* and in *you*.

In *Navy*, the situation is much more complex. Let us begin with the last sequence, since it is the only utterance: (25–26) *you / are in the ues navy now*. Here the illocutionary unit is the second one, the only one that cannot be cut out without compromising the interpretability of the utterance. In this case, it seems that more syllables are necessary to convey the functional focus (i.e., the illocutionary force); probably at least *ues navy now*. The other unit (*you*) is a Topic unit. We define Topic as the cognitive domain for the interpretation of the illocutionary force. When no Topic is present, the illocution is “unloaded” onto the context. We have found three different forms of Topic; all of them are found in Italian, Brazilian and European Portuguese, and in English (Cavalcante, 2016, 2018; Raso et al., 2017). In this text, there are two Topics with two different forms, the other one being *my friend* (18).

We have come now to the analysis of the three stanzas: (1–9), (10–17), and (18–24) respectively.

The first one is made up of seven intonation units (1–9) (plus two very small retracted units with one incomplete word), one of which is due to a fragmentation phenomenon (*the*): *so I went to this / &m [/1] what I thought was my friend / &he / &he[/1]this navy captain down at the / naval headquarters / I said / this is terribly awkward //*. This stanza features two patterns, the first encompassing units 1–7 and the second encompassing just the last two units. The illocutions of the two patterns are respectively *so I went to this / … this navy captain down at the / naval headquarters* and *this is terribly awkward*, in both cases the functional focus being on the last word. Notice that the first illocution is made up of three intonation units; in fact, after the first one, the illocution is interrupted by a parenthetic, and then a fragmentation phenomenon yields the perception of a boundary in the second part of the illocution. When an information unit is formed by two or more units, we call the units before the last one *scanning* (SCA) units, thus making it clear that they are part of a larger information unit. The parenthetic, a well-known unit in the literature (Schneider, 2007; Tucci, 2004), can be identified by its prosodic and functional features: different f0 profile (usually lower) from what precedes and what follows it, lower intensity and different articulation rate; functionally, the parenthetic provides comments about the locutive content of the pattern or part of it; distributionally, it can occupy almost any position, even inside another information unit, like in this case; the only position it cannot occupy is the absolute initial position of a TS. The second pattern, besides the illocution, features a Locutive Introducer (INT) (Giani, 2004; Maia Rocha & Raso, 2011). This unit basically has the function of introducing a meta-illocution, for example, as in this case, a reported speech. Its prosodic features are falling profile and higher articulation rate, as well as clear contrasts in f0, intensity and range with the units in reported speech.

The second pattern is made up of units 10–17, according to our segmentation: *I’ve just been promoted from / third mate / to second mate / and [/1] and / could we / possibly postpone these orders for a little bit //*. We can analyze this segmentation as the two following functional patterns: the first would be organized around two patterned illocutions (the second and the third units) preceded by a long INT. The second one would be realized by units 14–17. Here a fragmentation phenomenon occurs (*and [/1] and /*). The second *and* can be analyzed in two ways: either as a scanning unit or a unit called Discourse Connector (DCT). This unit has been described as a unit that connects different sub patterns in *stanza* or that begins utterances by marking continuity with the previous one (Frosali, 2008; Raso, 2014); nevertheless, no specific prosodic features have been found so far for DCTs, apart from their tendency to exhibit long duration (Raso & Ferrari, forthcoming). Therefore, we still need more research to understand whether DCT really is a specific IU or should be treated as a SCA. But this is a minor issue for our purpose. If we do not distinguish between SCA and DCT, the whole second pattern would be made up of three SCA IUs composing one single illocutionary unit.

There are different possible segmentations that deserve to be mentioned. The first pattern could be *I've just been promoted from third mate / …* This alternative was chosen by three of the six expert annotators; as for the other three experts, two put a boundary after *from* and the other put a boundary between *promoted* and *from* (showing that he perceives a boundary approximately in the same position). With this latter segmentation, we could interpret the first unit (*I've just been promoted from*) as a Locutive Introducer that announces an illocutionary pattern (CMM), which is sort of isolated so as to be emphasized. Our measurements support the choice for this second alternative.

The second pattern could be *could we / possibly postpone these orders / for a little bit //*. This was the preferred solution among the 16 annotators (11 vs. 5), but not among the six experts (2 vs. 4). Segmenting as the majority did, we would have a unit of Appendix of Comment (*for a little bit*), that is, a unit that integrates the text of the illocution. The prosodic parameters of this unit are basically a falling profile and low intensity. Our measurements do not actually support the placement of a boundary after *orders*; moreover, *for a little bit* can easily be interpreted as a coda (i.e., a post-nuclear part) of the illocution. From a functional point of view, the two interpretations are not significantly different.

The third stanza (18–24) features a Topic in the first unit (*my friend*), three illocutionary juxtaposed units (*stood up / behind his desk* and *in his full &f*[/1] *four stripes*), the third one being made up of two units (there is a SCA boundary probably due to a fragmentation phenomenon), and a last pattern formed by a Locutive Introducer (*and said*) and a strong illocution of recall (*Lieutenant*) in reported speech (which continues in the last utterance).

5. Final remarks

We have conducted three different tasks on the two English texts. (1) we asked 16 annotators to segment the texts and observed the inter-rater agreement. Doing so, we differentiated within the 16 annotators and additionally considered the agreement among the six higher trained subjects. This allowed us to observe that, while the perception of boundaries is a very natural task, the distinction of their nature, whether terminal or non-terminal, is significantly improved by training; (2) we phonetically analyzed the boundary region, following what the literature considers as the major features responsible for conveying boundary perception. This allowed us to confirm the presence of the phonetic features mentioned in different phonetic studies where the annotators marked the boundaries, and at the same time to make a decision in a few cases of clear disagreement. This decision was made following phonetic findings, trying therefore to avoid theoretical motivations; (3) a functional analysis of the intonation units was implemented only after the segmentation, which was based on (1) and (2). Therefore, we tried to maintain the segmentation task independent from the task represented by the assignment of a functional value to the units that resulted from the segmentation process.

Acknowledgements

The authors would like to acknowledge CAPES, CNPq and FAPEMIG for the research grants that have made writing this chapter possible.

References

Barbosa, P. A. (2013). Semi-automatic and automatic tools for generating prosodic descriptors for prosody research. In B. Bigi & D. Hirst (Eds.), *TRASP 2013 Proceedings* (Vol. 13, pp. 86–89). Aix-en-Provence: Laboratoire Parole et Langage.

Barbosa, P. A., & Raso, T. (2018). Spontaneous speech segmentation: Functional and prosodic aspects with applications for automatic segmentation. *Revista de Estudos da Linguagem*, 26(4), 1361–1396.

Barth-Weingarten, D. (2016). *Intonation units revisited. Cesura in talk-in-interaction*. Amsterdam: John Benjamins. https://doi.org/10.1075/slsi.29

Bossaglia, G., Mello, H., & Raso, T. (this volume). Illocution as a unit of reference for spontaneous speech: An account for insubordinated adverbial clauses in Brazilian Portuguese. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Campbell, W. N. (1992). Syllable-based segmental duration. In G. Bailly, C. Benoıt, & T. R. Sawallis (Eds.), *Talking machines: Theories, models, and designs* (pp. 221–224). Amsterdam: North-Holland.

Cavalcante, F. (2016). The topic unit in spontaneous American English (Unpublished Master's thesis). Universidade Federal de Minas Gerais, Brazil.

Cavalcante, F. (2018). The Information Unit of Topic: A Crosslinguistic, Statistical Study Based on Spontaneous Speech Corpora (Unpublished PhD Dissertation). Universidade Federal de Minas Gerais, Brazil.

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

Cresti, E. (2011). The definition of focus in Language into Act Theory (L-AcT). In H. Mello, A. Panunzi, & T. Raso, *Pragmatics and prosody, illocution, modality, attitude, information patterning and speech annotation* (pp. 39–82). Firenze: Firenze University Press.

Cresti, E. (2018). The illocution-prosody relationship and the information pattern in spontaneous speech according to the Language into Act Theory. In M. Heinz & M. C. Moroni (Eds.), *Prosody: Grammar, information structure, interaction. Linguistik Online*, 88(1), 33–62. https://doi.org/10.13092/lo.88.4187

Cresti, E. (this volume). The pragmatic analysis of speech and its illocutionary classification according to Language into Act Theory. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Cresti, E., & Moneglia, M. (this volume). Some notes on the excerpts according to L-AcT. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Du Bois, J. W., Chafe, W. L., Meyer, C., & Thompson, S. A. (2000–2005). *Santa Barbara corpus of spoken American English, Part 1–4*. Philadelphia, PA: Linguistic Data Consortium.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76, 378–382. https://doi.org/10.1037/h0031619

Frosali, F. F. (2008). L'unità di informazione di ausilio dialogico: Valori percentuali, caratteri intonativi, lessicali e morfo-sintattici in un *corpus* di italiano parlato (C-ORAL-ROM). In E. Cresti (Ed.), *Prospettive nello studio del lessico italiano* (pp. 417–424). Firenze: Firenze University Press.

Giani, D. (2004). Una strategia di costruzione del testo parlato: L'introduttore locutivo. In F. A. Leoni, F. Cutugno, M. Pettorino, & R. Savy (Eds.), *Atti del convegno "Il parlato italiano" Napoli, 13–15.02 2003* (pp. 84–97). Napoli: M. D'"Auria.

t'Hart, J. (1981). Differential sensitivity to pitch distance, particularly in speech. *Journal of the Acoustical Society of America*, 69(3), 811–821. https://doi.org/10.1121/1.385592

Maia Rocha, B., & Raso, T. (2011). A unidade informacional de introdutor locutivo no português do Brasil: Uma primeira descrição baseada em *corpus*. *Domínios de Linguagem*, 5(1), 327–343.

Mittmann, M., & Barbosa, P. A. (2016). An automatic speech segmentation tool based on multiple acoustic parameters. *CHIMERA: Romance Corpora and Linguistic Studies*, 3(2), 133–147.

Moneglia, M. (2005). The C-ORAL-ROM resource. In E. Cresti & M. Moneglia (Eds.), *C-ORAL-ROM: Integrated reference corpora for spoken Romance languages* (pp. 1–70). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.15.03mon

Moneglia, M., & Raso, T. (2014). Notes on Language into act theory. In T. Raso & H. Mello (Eds.), *Spoken corpora and linguistic studies* (pp. 468–495). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61.15mon

Quené, H. (2007). On the just noticeable difference for tempo in speech. *Journal of Phonetics*, 35(3), 353–362. https://doi.org/10.1016/j.wocn.2006.09.001

Raso, T. (2014). Prosodic constraints for discourse markers. In T. Raso & H. Mello (Eds.), *Spoken corpora and linguistic studies* (pp. 412–467). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61.14ras

Raso, T., Cavalcante, F., & Mittmann, M. (2017). Prosodic forms of the topic information unit in a cross-linguistic perspective: A first survey. In A. De Meo & F. M. Dovetto, *La comunicazione parlata/Spoken communication* (pp. 473–498). Roma: Aracne.

Raso, T., Ferrari, L. Uso dei Segnali Discorsivi in corpora di parlato spontaneo italiano e brasiliano. In: Ferroni, R., Birello, M. (forthcoming). *La competenza discorsiva a lezione di lingua straniera*. Roma: Aracne.

Raso, T., Mittmann, M., & Oliveira Mendes, A. C. (2015). O papel da pausa na segmentação prosódica de corpora de fala. *Revista de Estudos da Linguagem*, 23(3), 883–922. https://doi.org/10.17851/2237-2083.23.3.883-922

Raso, T., & Rocha, B. (2017). Illocution and attitude: On the complex interaction between prosody and pragmatic parameters. *Journal of Speech Science*, 5, 5–27.

Rocha, B., & Raso, T. (2016). The interaction between illocution and attitude, and its consequences for the empirical study of illocutions. In C. Bardel & A. De Meo (Eds.), *Parler les langues romanes: Proceedings of the international GSCP conference (Stockholm, 2014)* (pp. 69–88). Napoli: Università degli Studi L'Orientale.

Schneider, S. (2007). *Reduced parenthetical clauses as mitigators: A corpus study of spoken French, Italian and Spanish*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.27

Senn, P., Kompis, M., Vischer, M., & Haeusler, R. (2005). Minimum audible angle, just noticeable interaural differences and speech intelligibility with bilateral cochlear implants using clinical speech processors. *Audiology and Neurotology*, 10, 342–352. https://doi.org/10.1159/000087351

Silber-Varod, V. (2011). The SpeeCHain perspective: Prosody-syntax interface in spontaneous spoken Hebrew (Unpublished doctoral dissertation). Tel- Aviv University, Israel.

Teixeira, B. H. (2018). Correlatos fonético-acústicos de fronteiras prosódicas na fala espontânea (Unpublished Master's thesis). Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.

Teixeira, B. H., Barbosa, P., & Raso, T. (2018). Automatic detection of prosodic boundaries in Brazilian Portuguese spontaneous speech. In A. Villavicencio, V. Moreira, A. Abad, H. Caseli, P. Gamallo, C. Ramisch, H. G. Oliveira, & G. H. Paetzold (Eds.), *Computational processing of the Portuguese language* (pp. 429–437). New York, NY: Springer.

Teixeira, B. H., & Mittmann, M. M. (2018). Acoustic models for the automatic identification of prosodic boundaries in spontaneous speech. *Revista de Estudos da Linguagem*, 26(4), 1455–1488.

Tucci, I. (2004). L'inciso: Caratteristiche morfosintattiche e intonative in un corpus di riferimento. In F. A. Leoni, F. Cutugno, M. Pettorino, & R. Savy (Eds.), *Atti del convegno "Il parlato italiano"* (CD-ROM -14 p.). Napoli: M. D'Auria.

Appendix A. Excerpt: *Hearts*

Rank	Speaker	IU and boundary type	Function	Dur	Art. rate		Pause	F0		Movement	Intensity			
				(smoothed z)	(syll/s)		(s)	(st)			(db)			
											global	*spectral emphasis*		
											means (diff)	*left*	*right*	*R-L*
1	DAN	wait //	COM	–2,14	–	–	–	1,1	–0,3	–	0,72	3,16	6,35	3,19
2	DAN	play novice /	CMM	–1,92	5,25	–	–	3,39	0,6	–	0,87	9	8,1	–0,9
3	DAN	I've never played hearts before <in my life> //	CMM	–2,19	6,55	6,94|7,81	–	–	–	–	–	–	–	–
4	JEN	<you've never> played hearts //	COM	–1,9	6,13	–	–	–	–	–	–	–	–	–
5	DAN	no /	COM	–0,8	–	–	–	–	0,01	fall-r	–2,01	7,23	8,02	0,79
6	DAN	I don't know how to play it //	COM	–2,31	8,36	9,91|8,16	–	–	–	–	–	–	–	–
7	JEN	oh //	COM	3,08	–	–	0,464	4,11	–0,83	–	0,78	30,61	20,12	–10,49
8	JEN	okay //	COM	–1	4,72	–	0,117	–1,97	–2,32	r-fall	4,74	27,83	24,26	–3,57
9	JEN	I'll teach you //	COM	–2,26	4,98	–	–	–	–	–	–	–	–	–
10	DAN	passing disabled /	CMM	–1,51	4,51	–	–	–	–	fall-r	–	–	–	–
11	DAN	<that's you> //	CMM	–1,73	3,48	–	–	–	–	–	–	–	–	–

Appendix B. Excerpt: *Navy*

Rank	IU and boundary type	Function	Dur	Art. rate		Pause	F0		Mov.	Intensity			
			(smoothed z)	(syll./s)		(s)	(st)			(dB)			
				entire unit	*first|final 4 syll.*		*points (diff)*	*means(diff)*		*global*	*spectral emphasis*		
							i.e. reset			*means (diff)*	*left*	*right*	*R-L*
1	so I went to this /	SCA	–0,2	6,98	–	0,408	3,86	0,39	–	–3,44	14,69	9,16	–5,53
2	&m [/]	EMP	0,34	0,00	–	–	0	–0,82	–	11,45	9,16	12,06	2,9
3	what I thought was my friend /	PAR	–0,26	5,21	6,10|4,57	–	–1,62	–3,06	–	–5,5	7,22	8,5	1,28
4	&he /	TMT	0,76	0,00	–	–	0,62	0,15	–	–6,09	8,49	12,17	3,68
5	&he [/1]	EMP	0,21	0,00	–	–	1,34	1,06	–	–1,19	12,17	4,85	–7,32
6	this navy captain down at the /	SCA	–1,19	5,35	4,98|5,53	–	2,39	1,48	–	–7,27	4,29	2,79	–1,5
7	naval headquarters /	COB	3,21	6,80	–	1,381	5,21	1,22	–	–3,69	13,73	6,14	–7,59
8	I said /	INT	–1,16	5,65	–	–	2,59	4,58	fall-r	1,18	13,32	9,32	–4
9	this is terribly awkward //	COM	–2,52	6,38	7,22|5,77	–	1,94	3,89	fall-r	5,02	15,57	13,44	–2,13
10	I've just been promoted from /	INT	–2,11	5,65	5,61|6,56	–	8,27	9,71	flat-r	27,31	3,19	15,1	11,91
11	third mate /	CMM	–1,26	3,94	–	0,615	–1,48	–2,89	–	–15,13	11,2	5,33	–5,87
12	to second mate /	CMM	0,27	7,52	–	–	2,43	1,83	–	–9,2	8,15	4,67	–3,48
13	and [/1]	EMP	–0,28	0,00	–	–	1,29	0,54	–	7,83	4,67	6,46	1,79
14	and /	SCA	–0,86	0,00	–	–	3,69	2,59	–	4,84	6,46	11,7	5,24
15	could we /	SCA	–1,28	7,04	–	–	0,83	0,92	–	–0,37	9,47	7,6	–1,87

Rank	IU and boundary type	Function	Dur	Art. rate		Pause	F0		Mov.	Intensity			
			(smoothed z)	(syll./s)		(s)	(st)			(dB)			
				entire unit	*first|final 4 syll.*		*points (diff)*	*means(diff)*		*global*	*spectral emphasis*		
							i.e. reset			*means (diff)*	*left*	*right*	*R-L*
16	possibly postpone these orders		–2,31	5,78	5,43|5,57	–	2,2	0,26	–	–5,54	10,8	6,64	–4,16
17	for a little bit //	COM	2,94	6,27	–	1,142	5,97	6,99	–	21,63	11,4	11,19	–0,21
18	my friend /	TOP	–1,36	4,69	–	–	1,93	–0,31	–	–1,77	9,06	12,53	3,47
19	stood up /	COB	–1,45	9,71	–	0,161	3,02	0,28	–	–7,83	16,12	10,61	–5,51
20	behind his desk /	COB	–0,45	4,72	–	0,264	2,81	1,48	–	–4,66	12,22	5,55	–6,67
21	in his full &f [/1]	SCA	0,78	3,35	–	–	–1,78	–3,54	–	–3,08	13,77	16,36	2,59
22	four stripes /	COB	–1,47	2,80	–	–	–0,32	–1	–	–0,41	15,54	2,94	–12,6
23	and said /	INT	0,57	4,26	–	0,462	6,92	7,24	–	17,39	11,13	7,94	–3,19
24	Lieutenant //	COM	0,67	5,60	–	0,462	–12,88	–7,98	–	–7,75	47,8	11,05	–36,75
25	you /	TOP	0,5	0,00	–	–	0	–5,55	–	–4,86	11,05	16,87	5,82
26	are in the ues Navy now //	COM	–1,2	4,85	5,41|5,05	–	–	–	–	–	–	–	–

CHAPTER 2

Analysis of two English spontaneous speech examples with the dependency incremental prosodic structure model

Philippe Martin
Université Paris 7 Denis Diderot UFRL, LLF

Two examples of English spontaneous speech are analyzed prosodically, using the dependency incremental prosodic structure model. Instead of annotating prosodic events with the ToBI system, stressed accent phrases and final syllables are described in terms of rising or falling melodic contours, characterized by their melodic change above or below the *glissando* threshold. These contours indicate dependency relations between accent phrases, which in turn define the sentence prosodic structure.

Keywords: dependency incremental prosodic structure model, accent phrases, stress group, melodic contours

1. Introduction

The analysis of the two English samples is based on the concept of Dependency Incremental Prosodic Structure (DIPS – Martin, 2009, 2015), which defines the prosodic structure as a hierarchical assembly of Accent Phrases (APs) elaborated incrementally along the time scale according to dependency relations instantiated by melodic contours placed on stressed syllables. As in the Autosegmental-Metrical approach, APs are defined as sequences of syllables with only one lexical stress carried by a content word (noun, verb, adverb, or adjective). Therefore, other types of syllabic stress, such as emphatic stress, are not part of this definition, although emphatic stress may occur and may even be located on a lexically stress syllable.

In non-lexically stress languages such as Korean or French, accent phrases (also called *stress groups*) end with a stressed syllable and can contain one or more lexical words, whereas in lexically stressed languages such as English or Italian, APs contain by definition only one lexical item (i.e., a content word), unless some de-accentuation took place in the stress group to leave only one syllable stressed

https://doi.org/10.1075/scl.94.11mar

in a group of content words. However, grammatical words can also be stressed in emphatic or slow speech styles.

It has been shown that stress groups (i.e., accent phrases) not only constitute the basic units of the prosodic structure, but also that they define the way the listener processes incrementally the oral linguistic information produced by the speaker. Indeed, stressed syllables carry, as an essential acoustic parameter, not only a longer duration (or some other distinctive acoustic pattern to differentiate them from unstressed syllables), but also a melodic movement, rise, fall, high, low, etc., which is not produced at random. Actually, these melodic movements encode a relation of dependency "to the right", that is, to some other stressed syllable belonging to another stress group occurring later in the sentence in order to form through an incremental process a hierarchy of accent phrases constituting the prosodic structure.

Each language or group of languages, such as Romance languages except French, share similar mechanisms to indicate the relations of dependency based on the contrast of melodic slope (Martin, 2015). To characterize prosodic dependency relations for English, I will rely on a somewhat old observation made of pairs such as *a **lea**der for a **chang**e* versus *a **lea**der for a **chang**e* (cf. Figure 1), respectively carrying either a high and rising melodic contour or a high and falling contour on its first stressed syllable ***lea**der*, which entrains a difference in meaning, that is, "a leader finally" and "a leader to change things".

This pattern is related to compound stress often described in the literature (Martin, 1977; Plag, 2006), although descriptions are often implying not melodic contours but presence or absence of stress. This is due to the common phonological belief that only a rising pitch can carry stress. Compound stress pertains to frequent constructions in English, such as *home **phon**e, **o**pera singer*, among others, characterized by a stress pattern described as 0–1 in the first case, and 1–0 in the second case. The presence of stress on the first component, associated with a rising pitch in the second case, contrasts with the first pattern, characterized by a so-called default stress and a falling pitch on the second component. A third pattern may be also considered, where the force of stressed syllables of both compounds are equal, leading to the pattern 1–1. In Figure 1, sentence terminal falling contours are labelled C0, non-terminal falling contours C2, and non-terminal rising contours belonging to a rise-fall pattern, C1.

As a rising contour is considered as marked, and a falling contour as unmarked, C2 (referring to a falling contour) of the first example (left of Figure 1) is neutralized. The phonetic differences between C2 and C0 are usually instantiated by either the slope of C0, steeper than C2, or the average height, lower for C0.

The melodic contours which indicate the prosodic structure are Cn, C1, C2, Cc, C0, whose phonological descriptions involve the concept of *glissando* (Rossi,

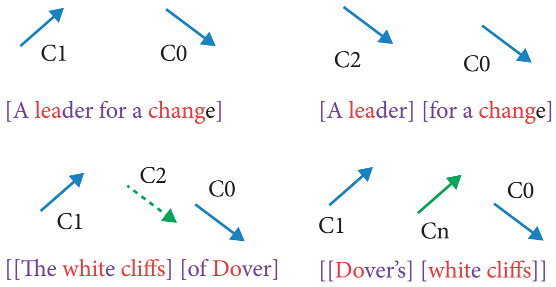

C1
C0
[A leader for a change]
C2
C0
[A leader] [for a change]
C1
C2
C0
[[The white cliffs] [of Dover]
C1
Cn
C0
[[Dover's] [white cliffs]]

Figure 1. Two characteristic pitch patterns for English

1971). As melodic contours do contrast essentially with melodic slope (i.e. rising vs. falling, or falling vs. rising), the perceptual aspect of this characteristics must be considered, even roughly, through the use of a *glissando* threshold. The other pertinent acoustic parameter is the melodic height, which may insure the necessary contrast between successive contours, with similar melodic slope, that is, rise-rise or fall-fall.

The phonological descriptions of C1 rising and C2 falling involve that their melodic change is above the threshold, so that their pitch movement is perceived as a rise or a fall. The neutralized contour Cn has a melodic variation below the *glissando* threshold, implying that its change in frequency is perceived as a static tone by listeners.

Cc is a complex contour, as it merges two distinct prosodic events, a somewhat flat contour on the accent phrase stressed vowel, below the *glissando* threshold, and a sharp rise on the last syllable, above the *glissando* threshold. If the stressed syllable is in final position, both melodic movements are combined on the syllable.

Many studies were conducted to derive some linguistic rule to predict which pattern will be used by speakers, be semantic, syntactic, phonologic, etc. However, recent research conducted on a very large number of compounds (Plag, 2006) concluded that there exists no clear correlation between the stress pattern used and any linguistic property of the compounds. The same speaker may use different realizations for the same compound in different conditions, which reinforce the hypothesis of a priori independence of prosody over any other structure existing in the sentence, as well as the precedence of the prosodic structure for both speakers and listeners (Martin, 2015). This latter hypothesis is supported by many examples found in spontaneous speech, where a chunk of a sentence prosodic structure planned and already started by the speaker would not accommodate the number of syllables of an intended syntactic pattern, forcing the speaker to either abandon the accent phrase in construction, or suddenly change its speech rate to fit the accent phrase duration (Martin, 2018).

An interesting case is related to the compound *toy factory*, which, according to the associated stress pattern, would mean "a factory manufacturing toys", with a stress pattern 1–0, or "a toy representing a factory", with a stress pattern 0–1 (Plag, 2006). With their associated pitch pattern, that is, rise-fall and [fall high]-[fall low], the two examples are illustrated Figure 2:

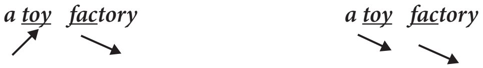


Figure 2. Melodic contour differences between the compound *a toy factory*

Sometimes, the pattern is hard to predict as in ***Oxford*** ***Street*** rise-fall compared to ***Oxford*** ***Road*** fall-fall. Martin (1977) suggested that the rise-fall pattern may be linked to some semantic "closeness" between components as estimated by the speaker at the time of enunciation. In a sense, *Oxford* and *Street* in the rise-fall realization may mean that the composing elements are conceptually closer that in *Oxford Road* with a fall-fall pattern.

Still two facts must be considered as well: (1) the possible neutralization of one of both stress syllables, and therefore of the associated pitch contour and (2) the stress clash condition which may force the elimination of one of the stress, depending on the duration left by the speaker between both potential stressed syllables. The duration of this interval is linked to the speech rate and to the structure of the syllables involved. For example, in ***House*** ***speaker*** versus *Vietnam* ***war***.

2. The dependency incremental prosodic structure

As seen above, the basic segmentation principle is based on the concept of stress group (Accent Phrase). In lexically stressed languages such as English, this definition leads to considered that an accent phrase contains obligatory one and only one "content word", such as a noun, a verb, an adverb or an adjective, as only these categories are supposed to have a lexical stress. Therefore, all other categories of "grammatical words", such as pronouns, auxiliaries, conjunctions, etc. may be part of an accent phrase but do not carry a stressed syllable.

However, analysis of spontaneous data shows that grammatical words can carry a stress, which may be therefore considered as an emphatic stress and not an accent phrase stress. Other difficulties with the definition come from possible de-accentuation of content words, implying the possibility to have more than one content word in a single accent phrase.

In any case, the accent phrase or stress group constitute the minimal prosodic unit in this analysis. The prosodic structure is then defined as the hierarchical classification of the sequence of accent phrases as encoded by the speaker. This hierarchy is indicated by specific prosodic markers, instantiated by vowel and syllable lengthening as well as the melodic contour located on (non-emphatic) stressed syllables, essentially on their vowels. As this process is dynamic in time, the prosodic structure is necessarily elaborated incrementally from local dependency relations "to the right", that is, from one prosodic marker to some other marker occurring later in the sequence of contours in the sentence.

For a specific language, the question is now to discover how the dependency relations between prosodic markers instantiated by melodic contours are encoded. Expected phonetic features possibly ensuring this function would be vowel duration, melodic height and rising or falling melodic variation, and vowel intensity (mostly for the terminal contour C0).

I will comment the two short excerpts along these theoretical lines, in an attempt to interpret the sequence of melodic contours placed on stressed syllables and ending large prosodic phrases, showing that pitch movements located on stressed syllables are not the fruit of hazard, but are there to indicate dependency relations and eventually the sentence prosodic structure intended by the speaker.

The starting point of the phonological description is the final terminal contour C0. Its acoustic parameters can vary, but this contour is normally falling to the lowest melodic height of the sentence. A perception test may ensure that listeners do not expect any continuation at this point, so that C0 indicates effectively the end of the sentence.

C1 and C2 are continuation rising (high) and falling (low) contours with a melodic variation above the *glissando* threshold. As shown above, the rising contour C1 is matched by a falling contour C2, usually placed at a lower melodic level.

Alternate configurations to C1–C2 are C1–C1 and C2–C2, with a contrast in height, the first contour being higher than the second.

3. *Navy* excerpt

The pitch and intensity curves displayed in the figures in this chapter were obtained with the WinPitch software package (Martin, 2003).[1] The validity of pitch curves was verified by comparing with a narrow band spectrogram eventually present on the figures. Every stress group is characterized by a highlighted pitch segment which determine the labelled melodic contour. The *glissando* threshold is computed automatically for each contour by the software.

[1] <https://www.winpitch.com/>.

Only remarkable pitch movements are discussed. The rest of the excerpt is most of the time neutralized with a fast speech rate. Most of the melodic patterns are based on the rise-fall pattern indicating a relation of prosodic dependency. This pattern can be embedded in another rise-fall pattern, as in Figure 3.

Stressed syllables are in bold and underlined, their corresponding melodic segments are highlighted.

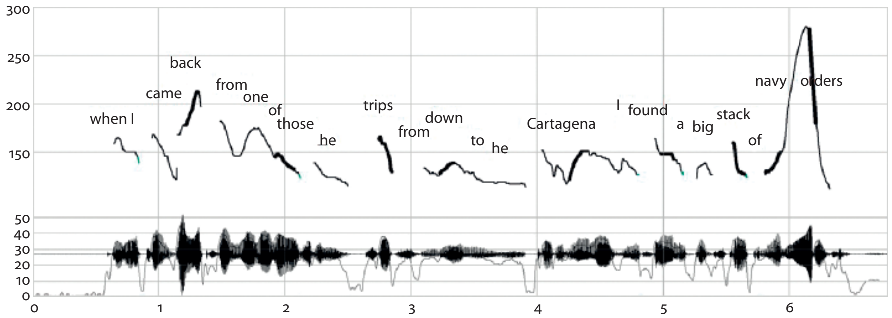


Figure 3. Pitch and intensity curves for [[*When I came* ***back*** C1 *from one of* ***tho****se* Cn *he* ***trips*** C2] *from* ***down*** Cn *to he Carta****ge****na* C1] [*I* ***found*** Cn *a* ***big*** ***stack*** Cn *of* ***na****vy* C1 ***or****ders* C0]

This first sentence shows a first embedding pattern [C1 Cn C2] in the sequence *When I came* ***back*** *C1 from one of* ***tho****se Cn he* ***trips*** *C2*. The same rise-fall pattern appears on ***big*** rise ***stack*** fall, but in this latter case the melodic movements are below the *glissando* threshold and are therefore labelled Cn (neutralized). Although a minor prosodic boundary has been perceived after *Cartagena*, the segment *from* ***down*** *Cn to he Carta****ge****na C1* presents a reduced melodic variation and may be interpreted as a prosodic parenthesis ended by a continuation contour C1. The last compound ***na****vy* ***or****ders* carry the same rise-fall pattern, but this time using a falling C0 instead of C2.

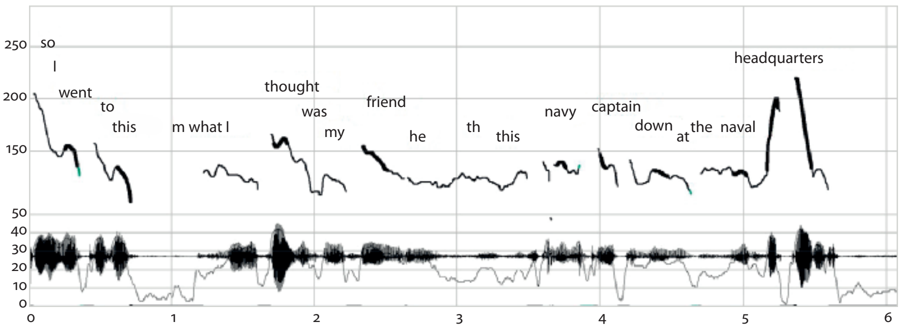


Figure 4. Pitch and intensity curves for [*so I* ***wen****t* Cn *to* ***this*** Cn *m* [*what I* ***thought*** Cn *was my* ***friend*** C0] *he th this* ***na****vy* Cn ***cap****tain* Cn *down at the naval* ***head*** C1 ***quar****ters* C0]

The sequence *I **thought** Cn was my **friend** C0* in Figure 4 is perceived as a complete well-formed prosodic structure, ended by a terminal conclusive contour C0 and defining a prosodic parenthesis. It can actually be removed with a sound editor resulting in the remaining sequence *so I **wen**t Cn to **this** Cn **na**vy Cn **cap**tain Cn down at the naval **head** C1 **quar**ters* C0 as both syntactically and prosodically well-formed. The last compound ***headquar**ters* carry the prototypic rise-fall pattern C1–C0 with a large melodic swing.

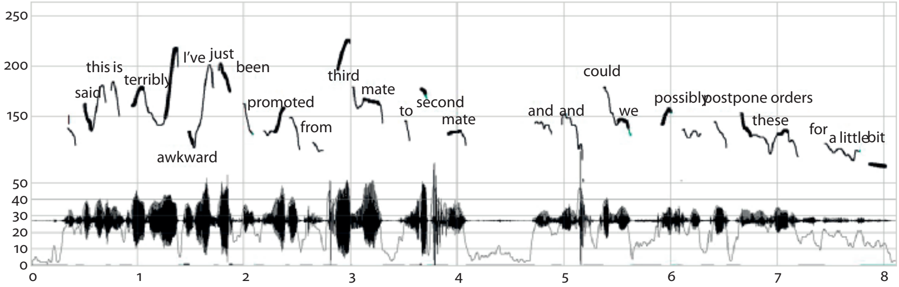


Figure 5. Pitch and intensity curves for [*I **said*** C2 *this is **ter**ribly* Cn ***aw*** C1 ***kward*** C0] [*I've **just*** Cn *been pro**mo**ted* C1 *from **third*** C1 ***mate*** Cn *to **se**cond* Cn ***mate*** C2] [*and and could we **po**ssibly* C1 *post**pon**e* C2 *these **or**ders* Cn *for a little **bit*** C0]

The segment *I **said** C2 this is **ter**ribly Cn **aw** C1 **kward** C0* in Figure 5 ends with a conclusive contour realized with a large melodic swing and a rise-fall pattern, each movement located respectively on the stressed and final syllables. The word *awkward* is prosodically divided in two syllables, carrying two specific melodic contours. Besides, the macrosyntactic boundary after *second mate* is marked with a falling C2 contour, ending the sequence C1 Cn Cn C2 in *from **third** C1 **mate** Cn to **se**cond Cn **mate** C2*. Again, the C1 C2 pattern is realized in *and and could we **po**ssibly C1 post**pon**e C2.*

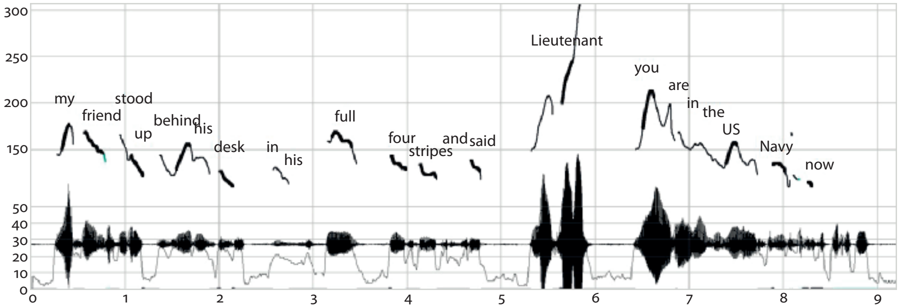


Figure 6. Pitch and intensity curves for [***my*** C1 ***friend*** C2 *stood **up*** C2 *be**hind*** C1 *his **desk*** C2] [*in his **full*** Cn ***four*** Cn ***stripes*** Cn *and **said*** C2] [*Lieu**te**nant* C1 ***you*** C1 *are in the **U**S* C1 ***Na**vy* C2 ***now*** C0]

In Figure 6, the sequence *__my__ C1 __friend__ C2 stood __up__ C2 be__hind__ C1 his __desk__ C2* present another example of successive rise-fall patterns, where consecutive falling contours C2 are differentiated by their melodic level, the second occurrence *stood __up__ C2* being lower that the first *__friend__ C2*.

We have also a remarkable imbedding of *__US__ C1 __Na__vy C2* inside *__you__ C1–C2 __now__ C0*. Again the pattern C1–C2 on *be__hind__ C1 his __desk__ C2* appears at the beginning of this section. The boundary after *stripes* carries only a Cn contour.

4. *Hearts* excerpt

In this second example, sentences are generally much shorter and frequently use the rise-fall melodic pattern. The accent phrase *before* in Figure 7 is buried in noise, but can be safely interpreted as a prosodic postnucleus, with a flat pitch contour. The accent phrase *in my life* is ended by a terminal conclusive contour C0, which can be observed on a narrow band spectrogram despite the other speaker voice overlapping. This segment appears as a “deferred complement”, it belongs to the sentence *I've never played __hearts__ C0 in my __lif__e C0*, divided prosodically by two independent prosodic structures.

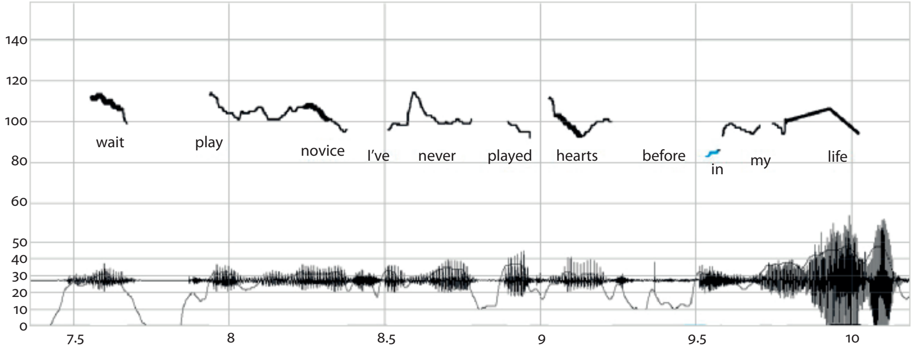


Figure 7. Pitch and intensity curves for [***wait*** C0] [*play __no__vice* C0][*I've never played* ***hearts*** C0] [*be__for__e* C0n] [*in my __lif__e* C0]

In Figure 8, there is a prosodically simple sequence, ended by a terminal interrogative contour C0i in *you've never __played__ Cn __hearts__ C0i*, and a sequence of short accent phrases ended by C0. In Figure 9, again, there is a short sequence with the second prosodic structure encoded by the C1 rising and C0 melodic contour.

In Figure 10, a slightly more complex two levels structure [*__pa__ssing* C1] [*di__sa__bled Cn that's __you__* C0] appears. The sentences of the second example are quite short, which implies the generation of simple prosodic structures with a limited set of contrast between melodic contours.

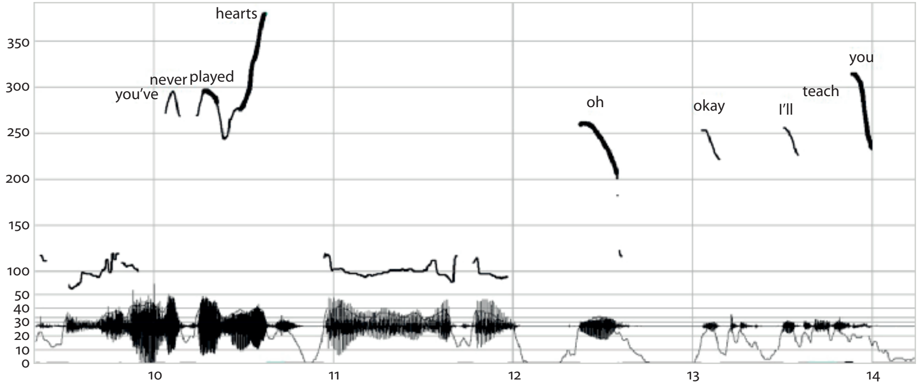


Figure 8. Pitch and intensity curves for [*you've never **played*** Cn ***<u>hearts</u>*** C0i] *(I don't know how to play it)* [***<u>oh</u>*** C0] [*o**<u>kay</u>*** C0] [*I'll teach **<u>you</u>*** C0]

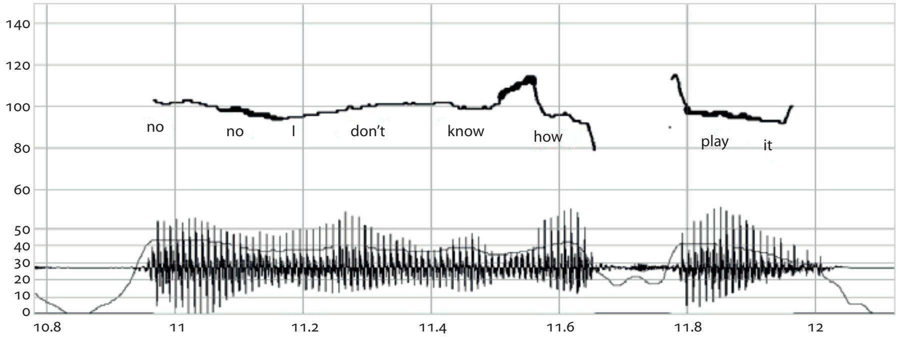


Figure 9. Pitch and intensity curves for [*No **<u>no</u>*** C0] [*I don't **<u>know</u>*** C1 *how to play it* C0]

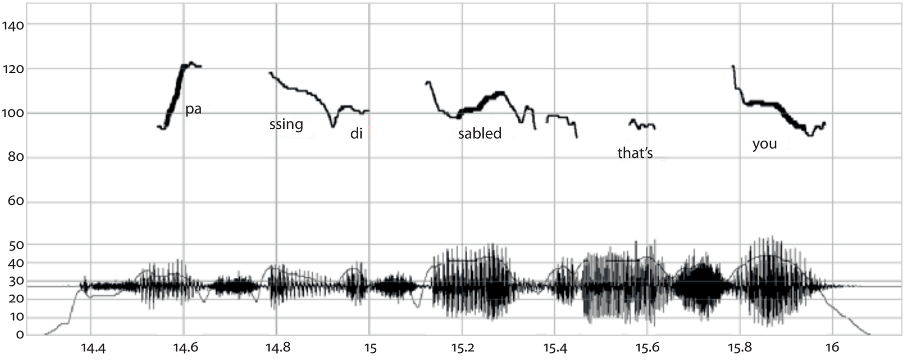


Figure 10. Pitch and intensity curves for [***<u>pa</u>**ssing* C1 *di**<u>sa</u>**bled* Cn *that's **<u>you</u>*** C0]

Conclusion

These two short excerpts show that melodic contours carried by stressed syllables are not realized at random. On the contrary, when their melodic variation is fast enough, that is, above the *glissando* threshold, they define a network of dependency relations between stress groups they belong, to ultimately define the completed prosodic structure intended by the speaker. This constitutes a basic mechanism for listener to recover the imbedded syntactic structure, which may or may not be congruent to the prosodic structure, that is, whose segmentation into accent phrases and whose hierarchical structure may not correspond to the syntactic structure.

References

Martin, P. (1977). A theory for English intonation. *Rapport d'Activités de l'Institut de Phonétique*, 11(1), 83–96.

Martin, P. (2003). Winpitch corpus, a software tool for alignment and analysis of large corpora. *Proceedings of the EMELD 2003*. Retrieved from <http://emeld.org/workshop/2003/martin-paper.pdf>

Martin, Philippe. (2009). *Intonation du français*. Paris: Armand Colin.

Martin, P. (2015). *The structure of spoken language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139566391

Martin, P. (2018). *Intonation, structure prosodique et ondes cérébrales*. London: ISTE.

Plag, I. (2006). The variability of compound stress in English: Structural, semantic, and analogical factors. *English Language and Linguistics*, 10(1), 143–172. https://doi.org/10.1017/S1360674306001821

Rossi, M. (1971). Le seuil de glissando ou seuil de perception des variations tonales pour la parole. *Phonetica*, 23, 1–33. https://doi.org/10.1159/000259328

CHAPTER 3

Applying criteria of spontaneous Hebrew speech segmentation to English

Shlomo Izre'el
Tel Aviv University

Taking prosody to be the leading component in speech segmentation, this chapter attempts to transfer segmentation methodologies from Hebrew to English spontaneous speech. Following a process of segmentation by perception of two English chunks a detailed acoustic analysis has been conducted, using acoustic criteria that have been found meaningful for similar analyses of Hebrew, as detailed in my chapter for Part I of this volume, "The Basic Unit of Spoken Language and the Interface Between Prosody, Discourse and Syntax: A View from Spontaneous Spoken Hebrew". This process has produced suggestive results. Further analysis into the interface of prosody with discourse has also been found meaningful. Some terminological issues are discussed as well.

Keywords: prosodic units, segmentation, spontaneous spoken language, Hebrew, English

1. Introduction

Speech segmentation is based first and foremost on the premises that prosody is a formal feature of spoken language, no less than segmental features; that prosody is the main tool we use for spoken language segmentation; and that for the recipient, prosody is the lead to perform a correct interpretation of the segmental structure and consequently a sound interpretation of the information conveyed. Furthermore, prosody, information structure and syntax integrate in spoken language structure, forming a coherent unity.

There seems to be a consensus among linguists that the fundamental, pivotal unit of spoken language is the *intonation unit*, being not only the primary unit of segmentation but also the unit of reference for dealing with analytical features of spoken discourse. Since it is useful to distinguish between the prosodic layer and the segmental layer of speech units, I prefer to use different terms for the two layers, as well as a third term for the combined speech stretch.

https://doi.org/10.1075/scl.94.12izr

A *prosodic module* (henceforth: PM; aka *intonation unit*) is the smallest prosodic unit that can be perceived by prosodic contours and prosodic boundaries. It can thus be regarded as the first-level unit of prosody relevant for the study of spoken discourse. The PM encapsulates a segmental unit of language to be termed *segmental module* (SM), forming together an *information module* (IM). The boundaries of either a SM or an IM are therefore defined by prosody. There are two main classes of boundaries: major (indicating terminality) or minor (indicating continuity). Both are indicated by their respective boundary tones. A major boundary is also the boundary of a *prosodic set*.

A *prosodic set* (PS) 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 ends in a major boundary, whereas any (optional) previous PM ends in a minor boundary. Whereas a PM encapsulates a *segmental unit* and forming together an *information module* (IM), a *prosodic set* encapsulates an *information set* or *utterance* (Utt).

In the chapter "The basic unit of spoken language and the interface between prosody, discourse and syntax: A view from spontaneous spoken Hebrew" (Izre'el, this volume, Part I), I have suggested that while the primary segmentation unit can indeed be the *prosodic module* (PM), a look at its interface with segmental features – as well as accounting for prosodic boundary criteria – points to the conclusion that the reference unit of spoken discourse is the *utterance* (Utt).

That work, which has been aimed at analyzing naturally occurring, spontaneous Hebrew, will be applied to two English excerpts as a comparative exercise initiated by Tommaso Raso and Alessandro Panunzi. In what follows, some notes on the segmentation and analysis of the two English excerpts, found in Appendices (A–C), are presented. I will first describe the prosodic cues for segmentation and annotation of boundaries (Section 2). This will be followed by comments on discourse annotation (Section 3) and syntactic annotation, along with some comments on the interface between prosody and syntax (Section 4). A few notes on individual units then follows (Section 5), before a concluding note on the benefits of such contrastive analysis for analyzing languages and by implication to the general study of language. As mentioned, Appendices B and C presents the analysis of the two English excerpts upon which this brief and surely preliminary study has been conducted.

2. Prosody: Segmentation and annotation

A *prosodic module* (PM) has been defined according to prosodic features, mostly intonational (= pitch related) ones, as consisting of a "single, coherent intonation contour" (Chafe, 1987, p. 22; Du Bois, Cumming, Schuetze-Coburn, & Paolino, 1992, p. 17). A coherent intonation contour, while rather easily perceivable, is hard to define in itself by acoustic, formal terms. Moreover, it is not easy to define a PM by any other internal criteria. In practice, segmentation of discourse flow into PMs is made by detecting their boundaries, whereas internal criteria are brought into consideration only secondarily (Cruttenden, 1997, Section 3.2; cf. also Chafe, 1994, pp. 57–60; Harrington & Cassidy, 1999, Section 4.6.4; among many others). This practice has been successfully used in transcribing large corpora (Cresti & Moneglia, 2005; Du Bois, 2004; Du Bois et al., 1992, 1993; cf. also Cheng, Greaves, & Warren, 2005, following the methodology of Brazil, 1997).

The segmentation of the two English excerpts offered here was carried out by perception, and the features listed are based on the premise that there are acoustic criteria upon which boundaries are perceived, among which the most prominent ones are: pitch reset, initial rush, final length, final tone (i.e., terminal f0 movement), and pause (Cruttenden, 1997, Section 3.2; Hirst & Di Cristo, 1998, pp. 35–36). The first four criteria have been successfully used in segmenting spontaneous Hebrew speech, with the following hierarchy found for Hebrew: final length > pitch reset > pause > initial rush (Amir, Silber-Varod, & Izre'el, 2004). These features have been detected by perception aided by acoustic observations in Praat <www.fon.hum.uva.nl/praat/>. They have been listed in the analytic tables following a notation on a perceivable complete intonation contour.

Pitch reset is the compared pitch level at the beginning of a PM with the final pitch level of the previous PM. It is, therefore, irrelevant (irr) at turn-initial position. Initial rush can be compared both IM-internally (i.e., to the rate of other syllables within the same IM), and IM-externally (i.e., to the length of the last syllable(s) of the previous IM). The latter is more important to account for boundary phenomena. Initial rush is noted as irrelevant when the number of syllables in the IM is too small (two or three; a single syllable is indicated by n/a) or when the IM begins with an accented syllable.

Final length refers to the length of the final syllable relative to previous syllables in the same unit or, in the case of a single-syllable IM, relative to neighboring IMs or the general rate of speech for this speaker in this context. For final length, measurements are also given, so that one can compare the actual duration of the final syllable and the total duration of the respective PM, giving notice to the number of syllables and internal suprasegmental structure, a task which I leave to experience

American-English linguists.[1] Still, perception has been used more than measurement in the annotation of both initial rush and final syllable length. The final tone (f0 movement) is the one observed for the last syllable: fall (f), rise (r), rise-fall (rf), level (l); where "level" indicates flat, slightly rising or slightly falling tone.

The columns representing the attestation of these features in the corresponding PMs are listed according to their relative order in the unit, following the first column that represents the perceived nature of the boundary in functional terms (major // or minor /) and the second which represents the existence of a perceivable complete "coherent" intonation contour.

3. Discourse annotation

At the discourse level, the actual function of the perceived boundary in the discourse stretch is noted in the first column of the discourse part: terminal (t) or continuing (c). Then follows the number of utterances (Utt) and the number of IMs within each utterance. The next column indicates the IM type: *substantive* (s), indicating that the IM contains "substantive ideas of events, states, or referents" or *regulatory* (r), indicating that the IM's function is to regulate interaction or information flow (Chafe, 1994, p. 63). There are also incomplete units, which Chafe (1994) termed *fragmentary* (f), that is, units that have not reached successful conclusion. I prefer to add an indication of the type of fragmentary IM as either substantive (fs) or regulatory (fr), the latter unattested in the analyzed excerpts. For Chafe, any unit which has not reached its successful conclusion in terms of information, would be regarded as *fragmentary*. I prefer to distinguish between units that have been truncated prosodically and units that are not perceived as prosodically truncated. The latter will be regarded – and accordingly termed – *suspended* IMs. Accordingly, the boundary of suspended units cannot be differentiated from any other continuing boundary, and have therefore been marked by the same symbol (/). After all, our analysis can lean only on the perception of the hearer, who would not know whether an uttered IM would be continued or whether the speaker would interrupt the information flow. In the latter case, the speaker may repeat the last uttered word(s), repair his previous speech, or discard the utterance altogether. The speaker may eventually return to it later or discard it for good. I take the term *suspension* to reflect these possibilities better than indicate such IMs as *fragmentary*. An example of a suspended IM is given in Navy_3 and perhaps also Navy_1 (see the comments on Navy_1–5 below). An example of a genuine fragmentary IM is Hearts_13.

[1] The given measurements are achieved manually by using Praat software.

4. Syntactic annotation

The last analytical level is syntactic. This level suggests that there is an interface between discourse units – *utterance*s – and syntactic units – *clause*s. As discussed in my chapter on Hebrew (Izre'el, this volume, Part I), I regard the *utterance* (Utt) to be the default domain of the *clause* (cl), whether a *clause* is encapsulated by a single IM or spreads over more than one IM. An Utt can further consist of more than a single clause; therefore, the Utt can be rather regarded as the default domain of a spoken sentence. An IM can thus consist of either a phrase, being a component of a clause; a clause; a clause extended by non-clausal elements (e.g., a vocative); or, more rarely, of a spoken sentence that consists of two or more clauses. The Utt is the biggest information unit that can contain a sentence. A sentence will not spread beyond the boundary of a single utterance. In other words, a major prosodic boundary indicates the terminal boundary of a sentence (and by implication also of a clause).

Nevertheless, there are cases where a syntactic relationship can be established between Utterances (cf. Mithun, 2002, 2005, 2008 for the extension of syntactic relations beyond sentences). There are two such instances among the analyzed excerpts. The IM Navy_15 is what is usually be regarded an "afterthought". Here it is noted as a *postnucleus* (pn), a term borrowed from Martin (2015) (see below). This Utt makes an extension to the Utt which ends in Navy_14. This latter Utt definitely shows the terminal point of the discourse domain for the fully fledged clause *could we possibly postpone this orders*. This latter IM ends in a falling terminal tone, thus indicating the end of the Utt and the encapsulated sentence (that includes a previous clause and a conjunction). The Utt Navy_15 is uttered in soft voice and flat intonation, noted here as *postfix* (pf; see below). It consists of no discourse or syntactic features that can serve as criteria for considering it a (fully fledged) clause. This analysis is similar to the macrosyntactic approach of the French school (Blanche-Benveniste, 2000, pp. 120–121, 2010, Section 4.4.3; Deulofeu, 2013; Martin, 2009, Section 4.2, 2015, Chapter 8). Another case, a more interesting one, is Hearts_4. This IM is another instance of afterthought, yet with a different intonation contour where prosodic prominence is carried by the word *life*. Macrosyntactic analysis usually terms this type of addition *suffix* (cf. the references above). Martin (2015, Chapter 8), separating between prosodic units and segmental units, suggests that *postnucleus* has two corresponding prosodic units: *postfix* and *suffix*, distinguished by their respective prosodic contour. Although both a *suffix* and a *postfix* have "well-formed" prosodic structures, a *postfix* is signaled by a "reduced melodic span" of its melodic contour. Taking this characteristic as a defining feature for a *postfix*, it definitely fits our IM Navy_15. As for Hearts_4, it would be defined as *suffix* in Martin's (2015) terms:

> Suffixes are well-formed prosodic structures placed after the Prosodic Nucleus. The only characteristic that differentiates them from a sequence of two independent prosodic structures associated with two successive utterances pertains to the syntactic dependency relation that must exist between the text segments associated with them and the text segment associated with the Prosodic Nucleus (Avanzi & Martin, 2007). Therefore, a Suffix is not a special kind of prosodic structure.
> (p. 222)

The meaningful prosodic contour observed in Hearts_4, being a modality signifier (Bally, 1965, para. 50; Martin, 2009, Chapter 1, 2015, pp. 68–71), may be taken as one of the criteria for establishing this IM as consisting a predicate and hence as a unipartite clause (i.e., one that does not include a subject; Izre’el, 2012, pp. 220–221, 2018b, Section 5). However, this option should be studied for (American) English and is far beyond the scope of the present paper and the proficiency of the present author. In any case, both these instances can be regarded as cases of syntactic relations across sentence boundaries and across utterance boundaries.

5. Comments on individual units

5.1 Hearts

Hearts_1: The Utt *wait* can be interpreted either as a substantive IM, referring to the actual game or as a regulatory unit, referring to the discourse flow.

Hearts_4: The rf tone on the final syllable is due to the unit accent (prominence). For the syntactic analysis see the discussion above.

Hearts_8,9: The two IMs consist of regulatory elements (the first is paralinguistic) and therefore not syntactically relevant.

Hearts_11: As noted in Appendix B, the tone on [bl̩d] <*bled*> seems to be a flat level tone, which does not reach the bottom of the pitch range. Moreover, the intensity level stays high. The tone on [ei] <*a*> is higher then on [bl̩d] <*bled*> due to the unit accent (prominence). The boundary can perhaps be perceived as major due to the lower tone of the final syllable. I, however, perceive it as a minor boundary and hence the interpretation of line 11–12 as a single Utt with two clauses. From the syntactic point of view, the clause in line 11 is unipartite, anchored in the extra-linguistic context (cf. Izre’el, 2018a, Section 4.1.2).[2]

[2] I thank Marianne Mithun for discussing Hearts_11–12 with me.

5.2 Navy

Navy_1–5: This sentence includes a matrix clause and an embedded clause (line 2). The disfluency phenomena at the beginning of IM2 and in IM3 may suggest that IM1 be a suspended unit. Still, the sentence as a whole seems coherent: There are two NPs in apposition (*this what I thought was my friend* and *this navy captain*).[3] Therefore, suspension is doubtful in this case. In any case, the alleged ambiguity seems to support the idea advocated above about the nature of suspension.

Navy_2: *Uh* is regarded as lengthening per se (Silber-Varod, 2011, 2013, Section 6.2.1.3). The preceding *friend* is 437 ms long, and together the duration is 653 ms.

Navy_3: Whether this stretch carries a coherent complete intonation contour remains to be determined. It seems to have none of the indicative acoustic features, still a boundary seems to be detected perceptually, although not without doubt.

Navy_7: The indicated final length (103 ms) is that of the sequence [əɹ] <*war*>. If one takes [kəɹ] <*kward*> to be the final syllable, the duration is 168 ms.

Navy_11: The (repeated) conjunction *and* serves here as a discourse marker (Schiffrin, 1987, Chapter 6).

Navy_15: The entire unit is uttered with low level pitch and fast. For the syntactic analysis see discussion above.

Navy_19–20: An alternative segmentation will regard the two PMs as a single one, in spite of the faster rate of Navy_20.

Navy_21: This IM is uttered in a loud voice. The boundary, with a rising tone gaining exclamative power, may alternatively be perceived as minor (continuing).

Navy_22: The rise-fall tone is due to prominence put on the topic, in addition to its occupying an IM on its own.

6. Conclusion

Taking prosody to be the leading component in speech segmentation, and following previous work on cross-linguistic prosodic units and corpora segmentation (Hirst & Di Cristo, 1998; Mettouchi, Vanhove, & Caubet, 2015; among others), the attempt to transfer segmentation methodologies from Hebrew to English spontaneous speech

[3] I thank Marianne Mithun for discussing this sentence with me.

has proved successful in many ways. Not only perception of units have managed to come up with meaningful speech units, but further analysis into the interface of prosody with discourse has been meaningful. Moreover, some glimpses into prima-facie “mismatches” between prosody and syntax (notably instances of “afterthought”) might find their solution by comparing analyses of similar cases in Hebrew. It is hoped that work along the lines suggested via similar contrastive analyses as offered in this section will enhance our understanding of spoken discourse structure.

References

Amir, N., Silber-Varod, V., & Izre’el, S. (2004). Characteristics of intonation unit boundaries in spontaneous spoken Hebrew: Perception and acoustic correlates. In B. Bel & I. Marlien (Eds.), *Speech prosody 2004* (pp. 677–680). Nara, Japan: ISCA.

Avanzi, M., & Martin, P. (2007). L’intonème conclusif: Une fin de phrase en soi? *Cahiers de Linguistique Française*, 28, 247–258.

Bally, C. (1965). *Linguistique générale et linguistique française.* Quatrième edition revue et corrigée. Berne: Éditions Francke.

Blanche-Benveniste, C. (2000). *Approches de la langue parlée en français*. Gap-Paris: Ophrys.

Blanche-Benveniste, C. (2010). *Le français: Usages de la langue parlée.* Avec la collaboration de Philippe Martin pour l’étude de la prosodie. Leuven: Peeters.

Brazil, D. (1997). *The communicative value of intonation in English*. Cambridge: Cambridge University Press.

Chafe, W. (1987). Cognitive constraints on information flow. In R. S. Tomlin (Ed.), *Coherence and grounding in discourse* (pp. 21–51). Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.11.03cha

Chafe, W. L. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago, IL: The University of Chicago Press.

Cheng, W., Greaves, C., & Warren, M. (2005). *A corpus-driven study of discourse intonation: The Hong Kong corpus of spoken English*. Amsterdam: John Benjamins.

Cresti, E., & Moneglia, M. (Eds.). (2005). *C-ORAL-ROM: Integrated reference corpora for spoken romance languages*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.15

Cruttenden, A. (1997). *Intonation* (2nd ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139166973

Deulofeu, J. (2013). Le rôle de l’élément *que* dans les phénomènes de subordination. In J.-M. Debaisieux (Ed.), *Analyses linguistiques sur corpus: Subordination et insubordination en français* (pp. 427–497). Cachan: Hermès & Lavoisier.

Du Bois, J. W., Cumming, S., Schuetze-Coburn, S., & Paolino, D. (1992). *Discourse transcription*. Santa Barbara, CA: Department of Linguistics, University of California.

Du Bois, J. W., Cumming, S., Schuetze-Coburn, S., & Paolino, D. (1993). Outline of discourse transcription. In J. A. Edwards & M. D. Lampert (Eds.), *Talking data: Transcription and coding in discourse research* (pp. 45–89). Hillsdale, NJ: Lawrence Erlbaum Associates.

Du Bois, J. W. (2004). *Representing discourse, Part 2: Appendices and projects*. Santa Barbara, CA: Linguistics Department, University of California.

Harrington, J., & Cassidy, S. (1999). *Techniques in speech acoustics*. Dodrecht: Kluwer. https://doi.org/10.1007/978-94-011-4657-9

Hirst, D., & Di Cristo, A. (Eds.). (1998). *Intonation systems: A survey of twenty languages*. Cambridge: Cambridge University Press.

Izre'el, S. (2012). Basic sentence structure: A view from spoken Israeli Hebrew. In S. Caddéo, M.-N. Roubaud, M. Rouquier, & F. Sabio (Eds.), *Penser les langues avec Claire Blanche-Benveniste* (pp. 215–227). Aix-en-Provence: Presses Universitaires de Provence.

Izre'el, S. (2018a). Unipartite clauses: A View from Spoken Israeli Hebrew. In M. Tosco (Ed.), *Afroasiatic: Data and Perspectives* (pp. 235–259). Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.339.13izr

Izre'el, S. (2018b). Syntax, Prosody, Discourse and Information Structure: The Case for Unipartite Clauses. A View from Spoken Israeli Hebrew. *Revista de Estudos da linguagem* 26/4, 1675–1726.

Izre'el, S. (this volume). The basic unit of language and the interface between prosody, discourse and syntax: A view from spontaneous spoken Hebrew. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Martin, P. (2009). *Intonation du français*. Paris: Armand Colin.

Martin, J.-P. (2015). *The structure of spoken language: Intonation in romance*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139566391

Mettouchi, A., Vanhove, M., & Caubet, D. (Eds.). (2015). *Corpus-based studies of lesser-described languages: The CorpAfroAs corpus of spoken AfroAsiatic languages*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.68

Mithun, M. (2002). Rhetorical nominalization in Barbareno Chumash. In L. Conathan & T. McFarland (Eds.), *Report #12: Proceedings of the 50th anniversary conference of the survey of California and other Indian languages* (pp. 55–63). Berkeley CA: University of California.

Mithun, M. (2005). On the assumption of the sentence as the basic unit of syntactic structure. In Z. Frajzyngier, A. Hodges, & D. S. Rood (Eds.), *Linguistic diversity and language theories* (pp. 169–183). Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.72.09mit

Mithun, M. (2008). The extension of dependency beyond the sentence. *Language*, 83, 69–119. https://doi.org/10.1353/lan.2008.0054

Schiffrin, D. (1987). *Discourse markers: Language, meaning and context*. In D. Shiffrin, D. Tannen, & H. Hamilton (Eds.), *The handbook of discourse analysis* 1 (pp. 54–75). Oxford: Blackwell. https://doi.org/10.1017/CBO9780511611841

Silber-Varod, V. (2011). The SpeeCHain perspective: Prosody-syntax interface in spontaneous spoken Hebrew (Unpublished doctoral dissertation). Tel-Aviv University, Israel. Retrieved from <http://www.openu.ac.il/Personal_sites/vered-silber-varod/download/Vered%20Silber-Varod%20Dissertation-7.pdf>

Silber-Varod, V. (2013). *The SpeeCHain perspective: Form and function of prosodic boundary tones in spontaneous spoken Hebrew*. Saarbrücken: LAP Lambert Academic Publishing.

Appendix A. Transcription conventions and abbreviations

In both Appendix B (*Hearts*) and Appendix C (*Navy*), each enumerated line holds a single *information module* (IM); *utterance*s (Utt) are separated by thicker lines; turns (in *Hearts*) are separated by double thicker lines. Turn numbers in *Hearts* are followed by their respective numbers in the original recording out of which this excerpt is extracted for in-depth analysis.

Prosody

//	major prosodic-set\utterance boundary
/	minor prosodic module\information module boundary
-	truncated word
–	truncated prosodic module\information module
f	fall (tone)
fr	fall-rise (tone)
l	level (tone)
pf	postfix
PM	prosodic module
pn	postnucleus
PS	prosodic set
r	rise (tone)
rf	rise-fall (tone)

Discourse

c	(prosodic boundary indicating) continuation
f	fragmentary information module
fs	fragmentary substantive information module
IM	information unit
IS	information set
r	regulatory information module
s	substantive information module
t	(prosodic boundary indicating) terminality
t\a	terminal\appeal (usually y\n questions; Du Bois et al., 1992, Section, 6.3, 1993, p. 55)
Utt	utterance

Syntax

{ }	embedded (clause)
cl	clause
conj	conjunction
intrj	interjection
neg	negation
suff	suffix
voc	vocative

Other

[ ]	overlap
<>	standard orthography transcription
irr	irrelevant
ms	milliseconds
n/a	not available
SM	segmental module

Appendix B. *Hearts*

line	turn	speaker	transcription		prosody										discourse				syntax
			segmental text	boundary	contour	preceding pause	pitch reset	initial rush	total duration (ms)	final syllable duration (ms)	final length	final tone	following pause	boundary function	speaker's Utt #	IM/Utt	IM type	Utt constituents
1	1(5)	DAN	wait	//	+	+	irr	n/a	332	332	+	rf	+	t	1	1	s ?	cl
2			play novice	//	+	+	—	irr	639	179	+	f	—	t	2	1	s	cl
3			I've never played hearts before	//	+	—	—	+	1170	267	+	f	—	t	3	1	s	cl
4			[in my life]	//	+	—	—	irr	500	239	+	rf	+	t	4	1	s	pn (cl)
5	2(6)	JEN	[you've never] played hearts	//	+	+	irr	—	1204	495	+	r	+	t\a	1	1	s	cl
6	3(7)	DAN	no	//	+	+	irr	n/a	274	274	+	rf	—	t	5	1	r	neg
7			I don't know how to play it	//	+	—	—	+	974	272	+	f	+	t	6	1	s	cl
8	4(8)	JEN	oh	//	+	+	irr	n/a	547	547	+	f	+	t	2	1	r	intrj
9			okay	//	+	+	+	irr	401	252	+	f	—	t	3	1	r	intrj
10			I'll teach you	//	+	—	+	irr	565	128	—	f	+	t	4	1	s	cl
11	5(9,11)	DAN	passing disabl[ed]	/ (//)	+	+	irr	—	1104	144	—	l	—	c	7	1	s	cl1
12			[that's you]	//	+	—	—	irr	665	342	+	rf	+	t	8	1	s	cl2
13	6(10)	JEN	[queen of sp-]	——	—	+	irr	n/a	543	n/a	—	n/a	+	n/a	5	1	fs	n/a

Appendix C. *Navy*

line	transcription		prosody										discourse				syntax
	segmental text	boundary	contour	preceding pause	pitch reset	initial rush	total duration (ms)	final syllable duration (ms)	final length	final tone	following pause	boundary function	Utt	IM	IM type	Utt constituents
1	so I went to this	/	+	+	irr	—	837	242	+	l	+	c	1	1	s	cl1 $\{\mathrm{cl}_e\}$
2	m- what I thought was my friend uh	/	+	+	+	+	1687	216	+	l	+	c		2	s	
3	n- the	/	(+)	—	—	irr	441	157	—	l	—	c		3	r	
4	n- this navy captain down at the	/	+	—	+	—	1515	134	—	l	—	c		4	s	
5	naval headquarters	/	+	—	+	+	852	204	—	l	+	c		5	s	
6	I said	/	+	+	+	irr	271	145	—	l	—	c		6	s	cl2
7	this is terribly awkward	//	+	—	+	—	994	103	—	f	—	t		7	s	cl3
8	I've just been promoted from	/	+	—	+	—	1194	165	—	l	—	c	2	1	s	cl1
9	third mate	/	+	—	+	irr	567	234	+	l	+	c		2	s	
10	to second mate	/	+	+	+	—	654	229	+	l	+	c		3	s	
11	and and	/	+	+	+	irr	612	390	+	l	+	c		4	r	conj
12	could we	/	+	+	+	irr	318	183	+	l	—	c		5	s	cl2
13	possibly	/	+	—	+	irr	449	110	—	l	—	c		6	s	
14	postpone these orders	//	+	—	+	—	1067	241	+	f	—	t		7	s	
15	for a little bit	//	+ (pf)	—	—	+	692	400	+	l	+	t	3	1	s	pn
16	my friend stood up	/	+	+	+	—	1109	279	—	l	—	c	4	1	s	cl1
17	behind his desk	/	+	—	+	—	1106	485	+	l	+	c		2	s	
18	in his full	/	+	+	+	irr	1108	525	+	l	—	c		3	s	
19	four stripes	/ (none)	+	—	—	irr	905	475	—	l	—	c		4	s	
20	and said	/	+	—	—	irr	453	360	+	l	+	c		5	s	cl2
21	lieutenant	// (/)	+	+	+	irr	637	178	—	r	+	t		6	s	voc
22	you	/	+	+	+	n/a	422	422	+	rf	—	c	5	1	s	cl3
23	are in the US Navy now	//	+	—	+	—	1637	244	—	f	+	t		2	s	

CHAPTER 4

Basic units of speech segmentation

Marianne Mithun
University of California, Santa Barbara

The segmentation of the monologue *Navy* and the dialogue *Hearts* described here is based solely on the acoustic signal. The unit of reference is the intonation unit as defined in the work of Chafe, characterized by a single, coherent pitch contour. Units defined by pitch often coincide with intensity, pauses, rhythm, and phonation type, though not always. In English they typically begin with a pitch reset followed by declination. Series of intonation units often form larger prosodic sentences, which can show an overall declination in pitch, often with intermediate pitch resets at the beginning of each unit. As shown by Chafe (1987, 1992, 1993, 1994, 1998, 2000, 2018), each unit tends to convey one new idea or focus of consciousness. They often correlate with syntactic constituents or sentences, though not always.

Keywords: intonation unit, prosodic phrase, pitch reset, declination, prosodic sentences

The segmentation of the recordings *Hearts* and *Navy* described here essentially follows the principles laid down by Chafe (1987, 1992, 1993, 1994, 1998, 2000, 2018) and underlying those discussed in Du Bois, Schuetze-Coburn, Cumming, & Paolino (1993).[1] It is carried out on the basis of the acoustic signal alone. The functions of the resulting units in speech can then be examined in a second phase of work.

The basic unit of reference is the intonation unit or prosodic phrase, generally defined most saliently by a single, coherent pitch contour. The units defined in terms of pitch coincide often but not always with several other kinds of features: intensity, pauses, rhythm (especially initial rush or acceleration, final lag or deceleration), and phonation type (such as creaky voice or vocal fry). Intonation units can show a variety of pitch patterns, and there is tremendous variation across languages, genres, and speakers.

[1] See tagset in Appendix.

https://doi.org/10.1075/scl.94.13mit

Often basic intonation units in English are characterized by an initial pitch reset followed by declination, an overall descent in f0, heard as pitch. An example can be seen in Figure 1. The pitch on the first word *You* was 212 Hz, and that on last word *now* ended at 114.2 Hz. Punctuation is used in transcription to reflect terminal contours rather than syntax. A period indicates a final contour, a kind of closure. There are a variety of final contours. They often though by no means always show a full fall in pitch. A comma indicates a continuing contour, often some kind of non-final fall. Capital letters or acute accents letters mark especially prominent syllables.

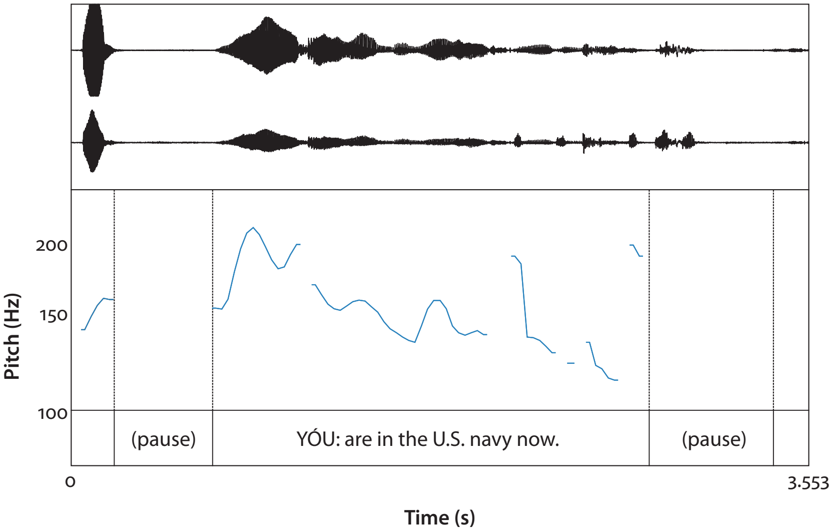


Figure 1. Pitch of basic intonation unit. *Navy* Turn 3: 00:00:30:12 – 00:00:32.073

Often there is also an overall fall in intensity, heard as volume, over an intonation unit. Figure 2 shows an intensity trace with a dotted line. Syllables with especially long duration can be marked with a colon: *YÓU:*.

Intonation units are often delimited by pauses. The intonation unit shown in Figures 1 and 2 was separated from the preceding unit by a pause of 0.444 s and followed by one of 0.785 s (during the second pause another participant was laughing).

Series of intonation units often form larger constructions called prosodic sentences, which can show an overall declination in pitch, often with intermediate pitch resets at the beginning of each unit. The pitch trace for a prosodic sentence from *Hearts* is in Figure 3. The pitch peak on the stressed syllable of *if you take tricks* is 283 Hz, that on *the highest card* is 249 Hz, that on *of the suit* is 219 Hz, and that on *takes the trick* is 187 Hz.

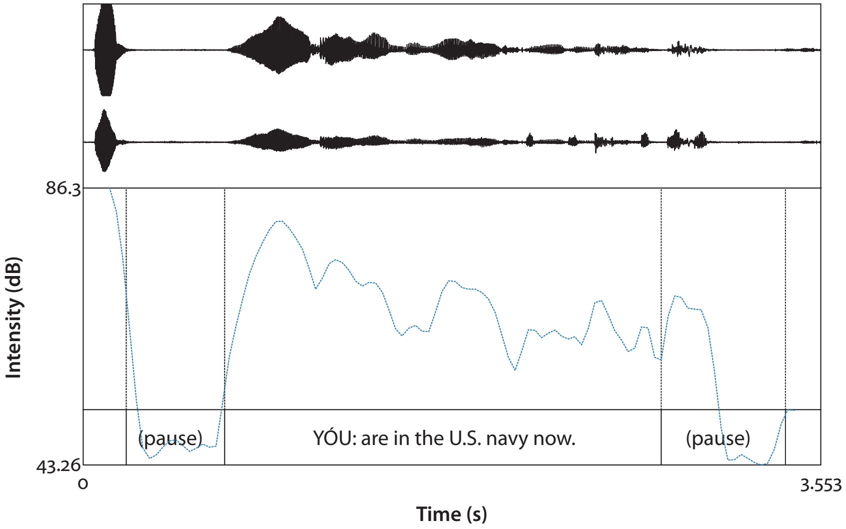


Figure 2. Intensity of basic intonation unit. *Navy* Turn 3: 00:00:30.012 – 00:00:32.073

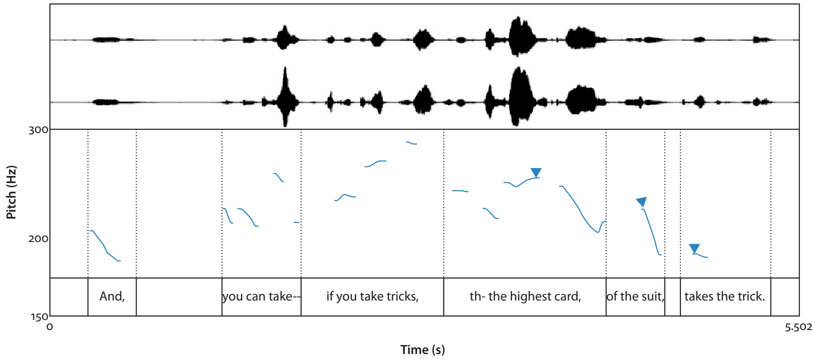


Figure 3. Prosodic sentence with declination. *Hearts* Turn 14: 00:00:35.610 – 00:00:40.610

Though many intonation units show similar declination in pitch, a variety of other patterns occur. Those in Figures 4 and 5 ended with a pitch rise.

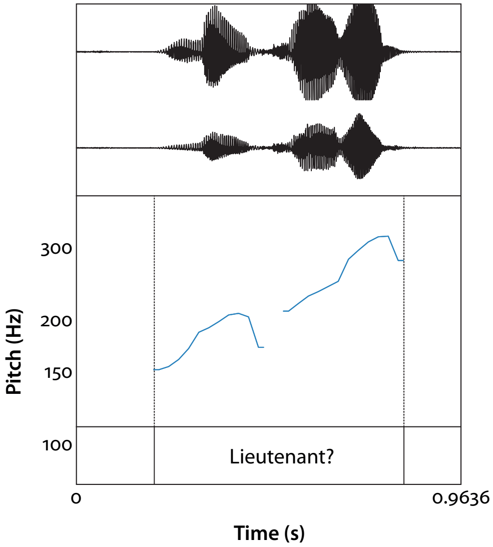


Figure 4. Terminal pitch rise. *Navy* Turn 3: 00:00:28.914 – 00:00:29.568

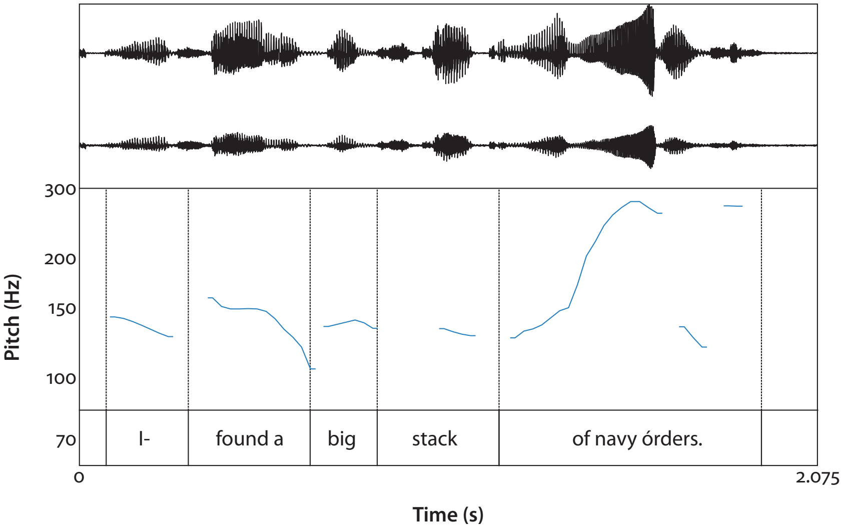


Figure 5. Terminal pitch rise. *Navy* Turn 1: 00:00:04.660 – 00:00:06.474

Because of the general pattern of declination, an intonation unit that does not show a significant fall in pitch can be perceived as rising, illustrated in Figure 6.

Figure 3 for *And, you can take-- if you take tricks, th-the highest card of the suit takes the trick*, shows another kind of intonation unit, truncation: *you can take--*. The speaker stopped, then began again with a full pitch reset. This contour

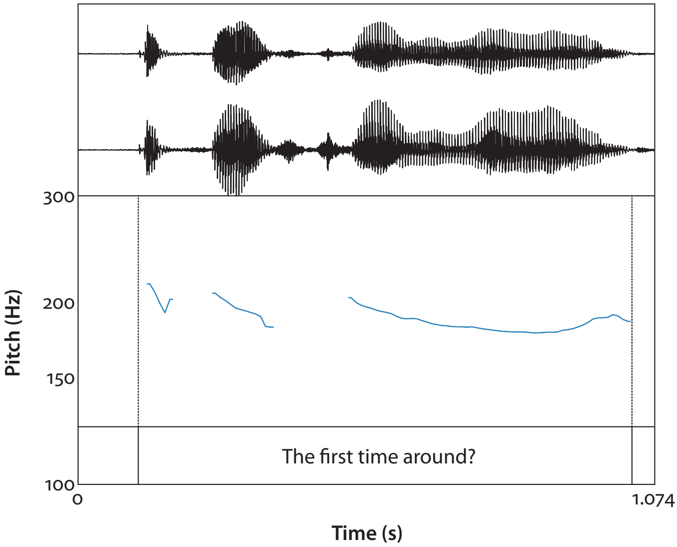


Figure 6. Perceived rise in the absence of expected pitch declination. *Hearts* Turn 22: 00:01:22.480 – 00:01:23.399

is indicated with a double hyphen or m-dash. Brief disfluencies, as in *th- the* are marked with a single hyphen or n-dash.

Many intonation units show an increase in intensity parallel to that in pitch. That seen above in Figures 1 and 2 shows parallel descents in pitch and intensity. That seen in Figure 5 with rising pitch shows a matching rise in intensity, indicated with the dotted line in Figure 7.

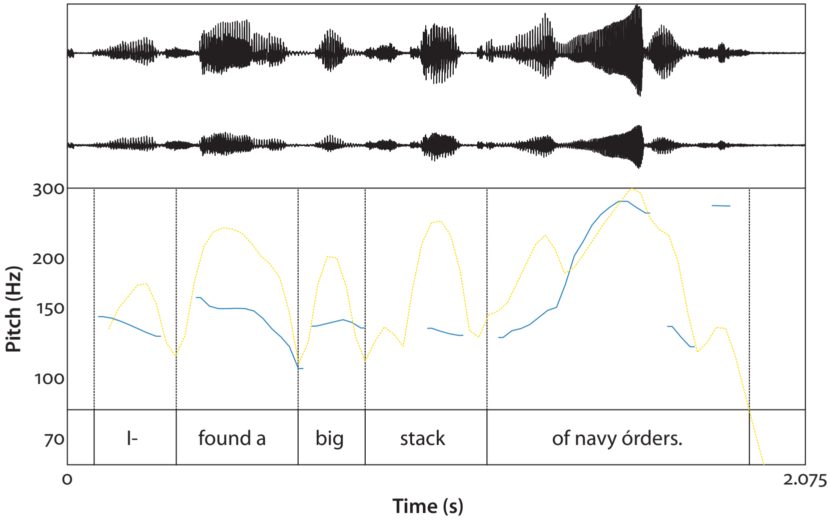


Figure 7. Parallel pitch and intensity rise. *Navy* Turn 1: 00:00:04.660 – 00:00:06.474

In the English sound files examined here, *Navy* and *Hearts*, the pitch and intensity generally operate in concert, though this is not always the case elsewhere.

Though many intonation units are separated by pauses, this is also not always the case. In the passage in Figure 8, each intonation unit begins with a pitch reset and shows declination, but there are no pauses.

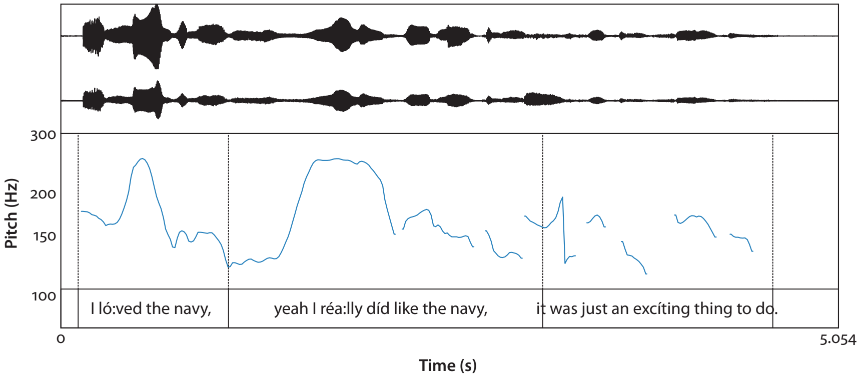


Figure 8. No pauses between intonation units. *Navy* Turns 11 and 13: 00:0:49.112 – 00:00:52.386

The variability of pausing can be seen in the passage in Table 1 from *Navy*. Each line represents a separate intonation unit. Numbers on the left identify the place of each unit in the *Navy* text. Immediately to the right of these identifications are the pause lengths in milliseconds that preceded each unit. There are clear divisions between units, but some are preceded by pauses, like 013 and 015, and some are not, like 014. Numbers at the far right on each line show the beginning and end times of that intonation unit in the *Navy* text.

Table 1. Variable pause times: Navy turn 3

013	1.38 ms	I said this is terribly áwkward,	00:00:15.205 – 00:00:16.434
014		I've just been promoted from,	00:00:16.434 – 00:00:17.633
015	0.05 ms	thírd mate,	00:00:17.683 – 00:00:18.204
016	0.15 ms	to sécond mate,	00:00:18.355 – 00:00:18.990
017	0.62 ms	and and--	00:00:19.609 – 00:00:20.185
018	0.04 ms	could we,	00:00:20.223 – 00:00:20.526
019	0.24 ms	possibly postpone these orders,	00:00:20.769 – 00:00:22.280
020		for a little bit.	00:00:22.280 – 00:00:22.721

In broad transcription in Chafe's system, two dots (..) indicate a brief pause, and three dots (…) a longer pause. Overlaps between speakers are indicated by brackets (see Table 2).

Table 2. Overlap: Hearts turns 3 and 4

006 Jen	0.39 ms	Wan[na play hearts?]	00:00:04.601 – 00:00:05.347
007 Dan		[Let's check that] one out.	00:00:04.796 – 00:00:05.677

Where there are multiple overlaps in close proximity, brackets can be marked with subscript numbers (see Table 3).

Table 3. Multiple overlaps: Hearts turns 17–19

053 Dan	0.44 ms	Rig[$_2$ht.]	00:00:51.984 00:00:52.225
054 Jen		[$_2$We] have three points in our hand [$_3$exactly.]	00:00:52.102 – 00:00:53.940
055 Dan		[$_3$And we w- wanna] try to get right of that.	00:00:53.392 – 00:00:54.745

Once a recording has been segmented into intonation units, the functions of the various patterns in speech can be investigated. Early on, Chafe (1987, and elsewhere) proposed that each intonation unit corresponds to one new idea, a single focus of consciousness. Some intonation units are regulatory (like *And* in Figure 3 *And, you can take--)*, some introduce a new referent, some introduce a new event or state, etc. Intonation units often correspond to syntactic constituents, and prosodic sentences to syntactic sentences, but not necessarily. The packaging of new ideas in intonation units can be seen in both of the transcripts. In the passage from *Navy* in Table 4, the speaker introduces each referent or characterization in a separate intonation unit.

Table 4. Navy, turn 3: One new idea at a time

009		So I went to this--	00:00:08.128- 00:00:08.931
010	0.39 ms	m- what I thought was my friend and and,	00:00:09.322 – 00:00:11.491
011		this navy captain down at the--	00:00:11.491 – 00:00:12.963
012		naval héadquarters.	00:00:12.963 – 00:00:13.872

Near the opening of *Hearts* (see Table 5), Jen first establishes the referent Hearts in one intonation unit, then characterizes it in another: *The card game*. Several turns later she incorporates this established referent into a larger sentence: *Wanna play hearts?*. In the next turn, Dan incorporates that concept into another longer unit: *I've never played hearts before*. (As can be seen from the intonation unit numbers, some material occurred between 003 and 006, and between 006 and 011.)

Table 5. Information packaging: Hearts, turns 1–2

002 Jen	Hearts.	00:00:01.509 – 00:00:02.069
003	The card game.	00:00:02.069 – 00:00:02.764
006	Wanna **play hearts**?	00:00:04.601 – 00:00:05.347
011 Dan	I've never **played hearts** before.	00:00:08.486 – 00:00:09.632

A similar exchange can be seen with the introduction of points (see Table 6), first with the light verb *have*. The idea is then enlarged with the number *three*, then enlarged again with *three points in our hand*.

Table 6. Information packaging: Hearts turns 15–18

044 Dan		and the queen of spaces,	00:00:45.757 – 00:00:47.027
045 Jen		are bad.	00:00:46.882 – 00:00:47.514
046 Dan		are the only thing.	00:00:47.027 – 00:00:48.146
047 Jen		that are	00:00:48.310 – 00:00:48.673
048 Dan		that--that have **points**.	00:00:48.568 – 00:00:49.589
049 Jen		that have **points**.	00:00:48.673 – 00:00:49.589
051	0.03 ms	Right.	00:00:49.618 – 00:00:49.943
052 Dan		So we got like--	00:00:49.943 – 00:00:50.457
053	0.12 ms	**three points right here**.	00:00:50.582 – 00:00:51.543
054	0.44 ms	Right.	00:00:51.984 – 00:00:52.225
055 Jen		We have **three points in our hand exactly.**	00:00:52.102 – 00:00:53.940

Intonation units or prosodic sentences often correlate with syntactic constituents or sentences, but not necessarily. When Dan explained that he had never played hearts before in his life, he introduced the information over two final intonation units, each with a final terminal contour (see Table 7). The second, which could simply be understood syntactically as an adverbial prepositional phrase, elaborated on the idea in the first.

Table 7. Information flow: Hearts turn 5

011	I've never played hearts before.	00:00:08.486 – 00:00:09.632
012	In my life.	00:00:09.632 – 00:00:10.161

As in many languages, a rising terminal contour often correlates with some kind of appeal. Here we can see that in the summons *Lieutenant?* in Figure 4 and the question *The first time around?* in Figure 6. In a number of other examples, rising pitch and intensity correlate with the focus or the most important information of the sentence, as in *I found a big stack of navy órders* in Figure 5. Heightened pitch, sometimes with corresponding intensity and/or duration, are often exploited for expressiveness. The heightened pitch of *really* can be seen in Figure 9.

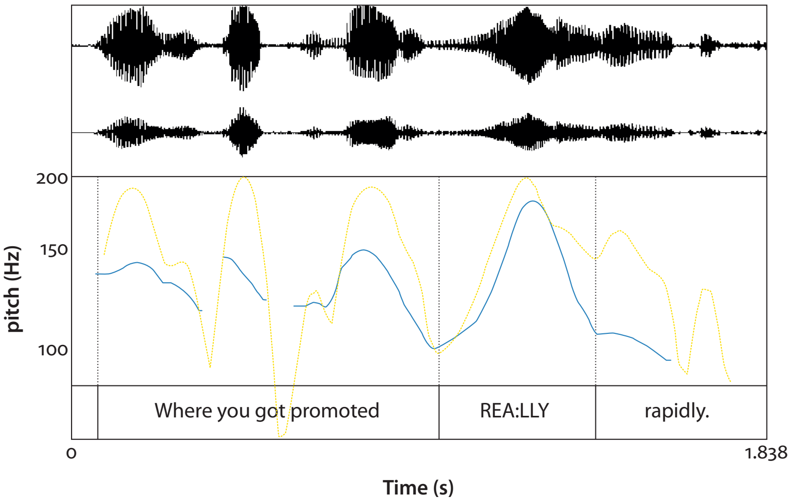


Figure 9. Heightened expressive pitch, intensity, and duration. *Navy* Turn 8: 00:00:39.356 – 00:00:41.277

Truncation, like that seen in Figure 3 *you can take-- you can take tricks*, typically indicates some kind of hesitation, often where the speaker is searching for an appropriate term or formulation. In couplet constructions a second intonation unit echoes the first prosodically, showing parallel pitch, intensity, and rhythm, rather than a continuing declination. The second unit reiterates information presented in the first or elaborates on it. Such a construction was seen in Figure 8 *I loved the navy; I really did love the navy*.

Both prosodic and syntactic constructions are conventionalized to a certain extent, and they often operate in concert, but they do not necessarily coincide, and they can vary across speakers and genres in different ways. The features that comprise prosodic patterns, primarily pitch contours but also intensity, timing, and pausing, vary along continua in ways that syntactic patterns do not, and can in many cases more directly reflect subtle semantic and discourse distinctions.

References

Chafe, W. (1987). Cognitive constraints on information flow. In R. Tomlin (Ed.), *Coherence and grounding in discourse* (pp. 21–51). Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.11.03cha

Chafe, W. (1992). Information flow. In W. Bright (Ed.), *Oxford international encyclopedia of linguistics* (Vol. 2, pp. 215–218). New York, NY: Oxford University Press.

Chafe, W. (1993). Prosodic and functional units of language. In J. A. Edwards & M. D. Lampert (Eds.), *Talking data: Transcription and coding in discourse research* (pp. 33–43). Hillsdale, NJ: Lawrence Erlbaum Associates.

Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago, IL: The University of Chicago Press.

Chafe, W. (1998). Language and the flow of thought. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 93–111). Mahwah, NJ: Lawrence Erlbaum Associates.

Chafe, W. (2000). The interplay of prosodic and segmental sounds in the expression of thoughts. In M. L. Juge & J. L. Moxley (Eds.), *Proceedings of the 23rd annual meeting of the Berkeley linguistics society*, 1997 (pp. 389–401). Berkeley, CA: BLS.

Chafe, W. (2018). *Thought-Based linguistics: How languages turn thoughts into sounds*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108367493

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of discourse transcription. In J. A. Edwards & M. D. Lampert (Eds.), *Talking data: Transcription and coding in discourse research* (pp. 45–89). Hillsdale, NJ: Lawrence Erlbaum Associates.

Appendix. Tagset

. (dot)	Full fall
?	Not specified
!	Not specified
, (comma)	Partial fall
…	Not specified
(blank)	Aligned unit without an explicit mark
-	Interruption (end of unit)
-L	Interruption (middle of unit)
--	Interruption (end of unit)
--L	Interruption (middle of unit)

CHAPTER 5

Segmentation of the English texts *Navy* and *Hearts* with SUU and LUU

Takehiko Maruyama
Senshu University / NINJAL

The chapter shows segmentation analyses of two English texts according to the criteria of the SUU (Short Utterance-Unit) and the LUU (Long Utterance-Unit). Our basic idea of segmentation is to identify utterance boundaries at two different levels. SUUs represent small information chunks, which are related to speakers' planning and hearers' understanding in a short time, and roughly correspond to prosodic and intonational units. LUUs, on the other hand, are basic chunks of interaction between the speaker and the hearer, corresponding to syntactic, discourse, and interactional units. Acoustic, prosodic, syntactic, and interactional boundaries were used as cues for segmenting utterances at two different levels. The technique of segmentation offers a way to view the multi-layered structure of spontaneous speech.

Keywords: SUU, LUU, prosodic/intonational boundaries, syntactic/interactional boundaries, multi-layered structure of spontaneous speech

1. Introduction: Two types of utterance units

In this section I will show analyses of segmenting the two English texts according to the criteria of the SUU, Short Utterance-Unit, and the LUU, Long Utterance-Unit (Den et al., 2010; Japanese Discourse Research Initiative, 2017; Maruyama, Den, & Koiso, this volume, Part I). The SUU and the LUU were originally designed as basic utterance units of Japanese dialogue. I will first give an overview of the SUU and the LUU, then examine how these units are applicable for segmenting the texts consisting of English monologue and dialogue.

Our basic idea of segmentation is to identify utterance boundaries at two different levels. The segments defined by these boundaries are called the SUU (Short Utterance-Unit) and the LUU (Long Utterance-Unit). The SUUs represent small information chunks which are related to speakers' planning and hearers' understanding in a short time, and they roughly correspond to prosodic and intonational

https://doi.org/10.1075/scl.94.14mar

units. Pauses of more than 0.1 s, or prosodic disjunctures, are identified as boundaries of SUU. The LUUs, on the other hand, are basic chunks of interaction between the speaker and the hearer, and they correspond to syntactic, discourse, and interactional units. LUU boundaries are identified at a final boundary of a main clause or a coordinate clause, which is normally realized with an intonation of final lowering. They are also identified at a point where an interaction between the speaker and the hearer occurs with tag questions, reactive tokens (backchannels, reactive expressions, and so on – see Clancy, Thompson, Suzuki, & Tao, 1996). Isolated filled pauses, word fragments and suspended utterances also constitute individual LUUs.

SUUs can be considered basic units of speakers' cognition and speech planning, while LUUs are characterized as basic units of syntactic chunk and participants' interaction. These two types are almost equivalent, respectively, to "Idea Units" proposed by Chafe (1994) and "Clausal Units" used by Biber, Johansson, Leech, Conrad, and Finegan (1999). Although they are not strictly hierarchical (see Maruyama et al., this volume, Part I), their boundaries can be regarded as segmentation points of a flow of utterance. In this section we regard both boundaries of SUU and LUU as utterance boundaries, and the segments defined by these boundaries are identified as reference units of spontaneous speech.

In the following sections I will examine some examples extracted from the English texts. The whole texts, segmented by the criterion of annotating SUU and LUU tags, are published online in the SLAC database. The transcription conventions are reported in the Appendix.

2. Some observations on the monologue *Navy*

Boundaries of SUU are identified in two ways: pauses of more than 0.1 s, and prosodic disjunctures, typically pitch reset. Figure 1 shows the string *I said this is terribly awkward I've just been promoted from third mate.*

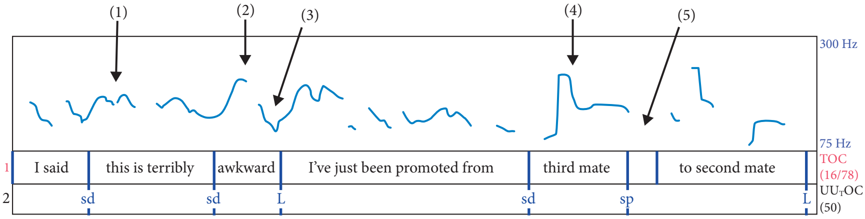


Figure 1. Annotating SUU and LUU annotation tags in the monologue *Navy*

The first part *I said this is terribly awkward* was divided into three parts with the tags **/sd** and /L, since pitch resets occur twice after *said* and *terribly*. Comparing the f0 peaks in each extent *I said, this is terribly* and *awkward*, the latter is higher than (or equal to) the former respectively (Figure 1; 1, 2), which means this utterance consists of three individual prosodic chunks. From the point of view of grammar and information structure, the first SUU is a main clause of the following quoted clause, the second is a topic and the third is a comment. *Awkward* ends with final lowering intonation (Figure 1; 3), which is distinct from natural declination; it shows that the utterance ends syntactically and phonologically at that point.

The second part *I've just been promoted from third mate to second mate* was also divided into three parts. The first break comes after *from* because a prosodic disjuncture exists there with a prominent pitch at *third* (Figure 1; 4). After *third mate* a pause of 0.143 s occurs (Figure 1; 5), which segments the utterance with the tag **/sp**.

Figure 2 shows two strings, *no one ever explained that to me before and a week later I was on my way out to Korea*, and *and then I was forced out because I failed a promotion to commander*, including clause linkages of coordinate clause and *because* clause.

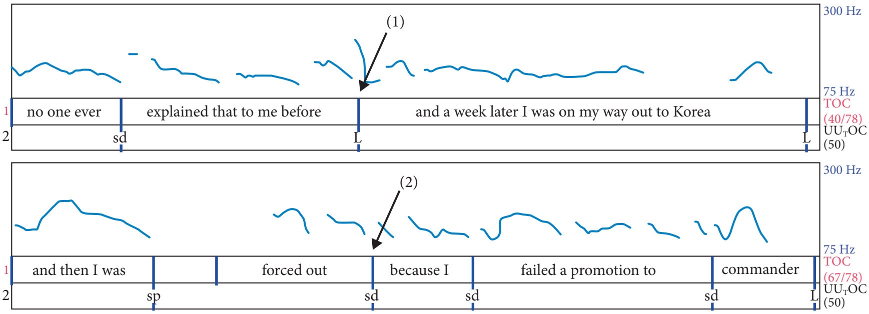


Figure 2. Clause linkages and identification of SUU and LUU

Boundaries of LUU are identified where an ongoing utterance ends with final lowering intonation and syntactic disjuncuture after a coordinate clause (Figure 2; 1). Subordinate clauses identified by, for example, *when* and *because*, also consist syntactic chunks. However, they are not regarded as LUUs, since they are dependent to other clauses. So, if a *because* clause follows a main clause, the point is not regarded as a syntactic disjuncture, but just as SUU boundary (Figure 2; 2). Coordinate clauses, on the other hand, form syntactically and semantically saturated chunks and their strings are identified as independent LUUs (Figure 2; 1).

Clancy et al. (1996) discussed reactive tokens, including expressions of backchannels, reactive expressions, collaborative finishes, repetitions and resumptive openers. Such expressions also form independent LUUs, like *hum hum, oh, no*, and so on. Tag questions *right* are used in the interaction from the speaker to the hearer, which also form independent LUUs as reactive tokens.

In (1) below, we report the transcription of the whole annotated turns containing the analyzed units, according to the format adopted in the online SLAC database. Turn numbering is here reported as in the original unannotated texts published in the SLAC database (see Introduction to Part II).

(1) *Navy*

3. *TOC: so I went to this **/L** m what I thought **/sd** was my friend &he th this navy captain down at the naval headquarters **/L** I said **/sd** this is terribly **/sd** awkward **/L** I've just been promoted from **/sd** third mate **/sp** to second mate **/L** and and could we possibly postpone these orders for a little bit **/L** my friend stood up **/sp** behind his desk **/sp** in his full f **/sp** four stripes **/L** and said **/sp** Lieutenant **/L** you are in the ues Navy now **/L** I <said **/sd** oh> **/L**
5. *TOC: no one ever **/sd** explained that to me before **/L** and a week later I was on my way out to Korea **/L**
15. *TOC: I stayed in the ues Navy **/sd** seventeen years and ten months **/L** and then I was **/sp** forced out **/sd** because I **/sd** failed a promotion to **/sd** commander **/L**

3. Some observations on the dialogue *Hearts*

In general the participants in a dialogue exchange short messages each other, and such a style can be observed in this dialogue *Hearts*. In (2), turns 1–3, a sequence *what's hearts, hearts it's the card game, oh yeah* is an example of an adjancy pair and a sequence closing third (Schegloff, 2007; Schegloff, Jefferson, & Sacks, 1977).

(2) *Hearts*

1. *DAN: what's hearts **/L**
2. *JEN: hearts **/R** it's the card game **/L**
3. *DAN: oh yeah **/R** put it up there **/L**

Thus, it is natural that we can observe reactive tokens much more frequently in dialogue than monologue. For example in turns 1–3 DAN asked *what's hearts* and JEN repeated the same noun *hearts*, which works as a reactive token by repetition.[1]

[1] Turn numbering is reported as in the original unannotated texts published in the SLAC database.

(3) *Hearts*
15. *DAN: okay /**R** so /**sp** h hearts and the queen of spades <are the only> thing /**sp** <that that have points> /**L**
16. *JEN: are <bad> /**R** that are points /**R** right /**R**

In (3), turns 15–16, while DAN's uttering *so h hearts and the queen of spades <are the only>*, right after the word *spades* JEN uttered *are <bad>* (Figure 3; 1). This is a reactive token as a collaborative finish, although after that DAN continued his utterance *<are the only>* so they overlapped. Finally, DAN and JEN conclude their utterances with almost the same expressions, *that that have points* and *that are that are points* simultaneously (Figure 3; 2). Figure 3 shows the extent.

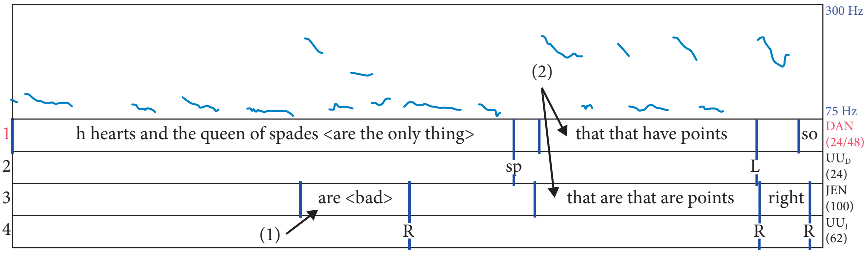


Figure 3. Collaborative finish in the dialogue

(4) *Hearts*
21. *DAN: [...] why is that /**L**
20. *JEN: [...] and these are two high /**L**
22. *JEN: why /**R** just because and cause /**sp** you should always pass a club /**sp** so that the person /**F** so the first hand /**F** everyone has a club so that they can't /**sp** discard a heart /**sd** cause you always assume that everyone's t /**F** no one is void of a suit /**sp** the first time around /**L** so you don't have to worry about throwing a high <card> /**L** and then I'm gonna throw two high cards /**L**

In the case of turns 20–22 in (4), DAN murmured *why is that* and JEN said *and these are two high* /**L** *why* /**R** *just because and cause* /**sp**.[2] At first JEN continued her explanation, after that she partly repeated DAN's question *why* and started her answer by *because*. This is also an example of a reactive token with repetition.

Reactive expressions are one reactive tokens with short lexical words or phrases frequently observed like *oh yeah, okay* and *right*. But some of those are

[2] Turns 20 and 21 are partially overlapped. Since the final part of turn 21 ends before the turn 20, we reported them in our example following the linear sequence of the uttered units.

not reactions to the preceding utterance (pre-sequence) but for the pre-action on the computer screen.

(5) *Hearts*

12. *JEN: &he first lead rotates **/L** first **/L** yeah **/R** always pass left **/L** alright **/R** […]

For example in (5), turn 12, JEN uttered *first lead rotates* **/L** *first* **/L** *yeah* **/R** *always pass left* **/L** *alright* **/R**, these reactive tokens *yeah* and *alright* are not reactions for DAN's utterance but for his action on playing hearts.

4. Discussion

Participants generally exchange short messages in a dialogue. However, in this dialogue JEN frequently generated narratives. This is because this dialogue is one which JEN instructs DAN how to play hearts, and when she explains a series of procedures it makes a kind of storytelling. Figure 4 shows an example.

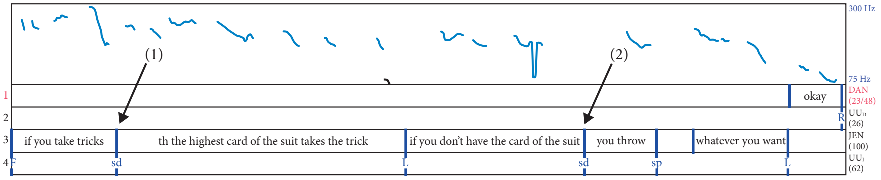


Figure 4. A narrative sequence in the dialogue *Hearts*

Figure 4 shows prosodic disjunctures occur between *if* clauses and main clauses (Figure 4; 1,2). JEN explained the procedure of playing hearts with two *if* clauses, and prosodic disjuncture occurred after them. The **/sd** tags seem to appear only in such narrative circumstances in the whole dialogue. In the monologue, on the other hand, a speaker basically keeps on speaking for a long time. The number of **/sd** tags annotated to the monologue *Navy* is much higher than that of the dialogue *Hearts*. This is because a concatenation of intonation phrases and clause linkage structures is frequently produced when speakers recount a series of episodes or explain procedures in narrative contexts. The distribution of annotated tags in the monologue and the dialogue is shown in Figure 5.

The ratios of LUU (R, F, L) and SUU (sp, sd) are significantly different between the two texts. SUUs in the monologue comprise 24 units out of 65 in total (36.9%), while in the dialogue they make up only 15 out of 88 in total (17.0%). In the monologue the speaker often relates an extended series of episodes, which produces

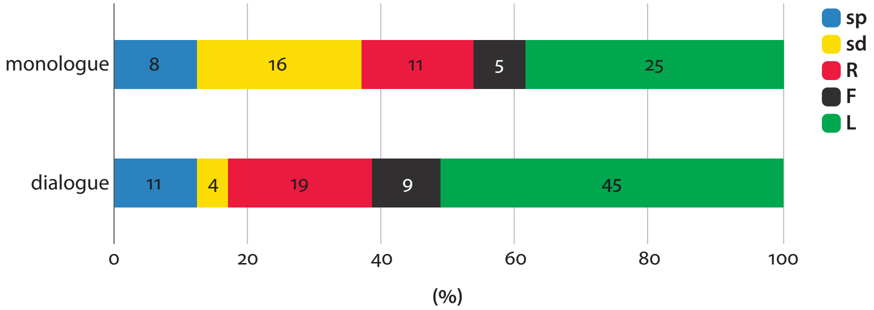


Figure 5. Distribution of annotated tags

a narrative with complicated structure. In such a situation it is natural for more pauses of more than 0.1 s and prosodic breaks to appear.

In the dialogue, on the other hand, the participants exchange short messages with each other. Each message consists of an LUU, which results in the ratio of LUUs being much lower than that for monologue (38.5% in dialogue compared to 51.1% in monologue). Especially in this text the participants speak rather quickly without pauses or prosodic breaks, which is reflected in the difference of ratios of **/sd**: only 4 (4.5%) in dialogue compared to 16 (24.6%) in monologue.

Examining the reactive tokens (/**R**), the inventories of expressions are different in the two texts. *Hum hum* and *right* account for 63.6% of reactive tokens in monologue, utilized by the listeners to show their agreement. On the other hand, *yeah, right, okay, exactly* and *alright* account for 63.2% of reactive tokens in dialogue, uttered by the both of participants in the conversation.

5. Concluding remarks

In this analysis, acoustic, prosodic, syntactic, and interactional disjunctures were used as cues for segmenting utterances on two different levels. The results of segmenting English monologue and dialogue texts into SUU and LUU were then analyzed. The technique offers a way to view the multi-layered structure of spontaneous speech.

Since the accent system is different between Japanese and English, a further issue is to examine how SUU and LUU may or may not be applicable to the characterization of English utterances more precisely, using larger speech data.

References

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

Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago, IL: Chicago University Press.

Clancy, P. M., Thompson, S. A., Suzuki, R., & Tao, H. (1996). The conversational use of reactive tokens in English, Japanese and Mandarin. *Journal of Pragmatics*, 26, 355–387. https://doi.org/10.1016/0378-2166(95)00036-4

Den, Y., Koiso, H., Maruyama, T., Maekawa, K., Takanashi, K., Enomoto, M., & Yoshida, N. (2010). Two-level annotation of utterance-units in Japanese dialogs: An empirically emerged scheme. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, & D. Tapias (Eds.), *Proceedings of the 7th language resources and evaluation conference (LREC2010)* (pp. 2103–2110). Valetta, Malta: European Language Resources Association (ELRA).

Japanese Discourse Research Initiative. (2017). *Utterance-Unit Labeling Manual* (Version 2.1). Retrieved from <http://www.jdri.org/resources/manuals/uu-doc-2.1.pdf>

Maruyama, T., Den, Y., & Koiso, H. (this volume). Design and annotation of two-level utterance units in Japanese. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis* (Vol. 1). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511791208

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53(2), 361–382. https://doi.org/10.1353/lan.1977.0041

Appendix. Tagset

/sp	Boundary of SUU with a pause more than 0.1 s
/sd	Boundary of SUU with a prosodic disjuncture
/R	Boundary of LUU with a reactive token
/F	Boundary of LUU with a fragment or suspended utterance
/L	Boundary of LUU with syntactic/interactional cue

If an utterance ends syntactically followed by a long pause, both of **/sp** and **/L** can be annotated, then the latter has priority.

CHAPTER 6

The Moscow approach to local discourse structure

An application to English

Andrej A. Kibrik[i], Nikolay A. Korotaev[ii] and Vera I. Podlesskaya[ii]
[i]Institute of Linguistics RAS and Lomonosov Moscow State University /
[ii]Russian State University for the Humanities

This chapter is an exploratory study in which we apply an approach to local discourse structure and prosody, developed for spoken Russian, to English talk. A key conceptual element of our approach is the notion of elementary discourse unit (EDU). EDUs are identified on the basis of prosodic criteria and demonstrate substantial correspondence to clauses. A range of structural, prosodic and discourse-semantic phenomena are reviewed, including pausing, discourse accent, phase, and spoken sentence. The analysis begins with those phenomena that are characteristic of both monologic and multi-party discourse, and proceeds with those features that are only found in interactional exchange. The Russian-oriented system of discourse transcription and analysis turns out to be generally applicable to the English evidence.

Keywords: spoken discourse, discourse transcription, local discourse structure, elementary discourse unit, prosody, pause, discourse accent, phase, spoken sentence

1. Introduction

In this chapter we attempt to apply the approach we had developed for analysis of spoken Russian discourse to English evidence. A brief sketch of the approach is provided here, for the reader's convenience, while further details can be found in our paper about Russian (Kibrik, Korotaev, & Podlesskaya, this volume, Part I); see also Kibrik and Podlesskaya (2009; in Russian), Kibrik (2011) and Fedorova & Kibrik (2020).

Local discourse structure consists of those units that can be considered minimal, or elementary, steps of discourse production. Linearly, discourse is organized as a sequence of steps (or quanta, pulses, spurts, etc.). Segmentation into quanta is

https://doi.org/10.1075/scl.94.15kib

both a theoretical and a practical issue. Theoretically, it is of interest to understand what is the size of the quanta, why the size is the way it is, how the boundaries between the quanta are established, and so forth. From a practical point of view, criteria are needed allowing a transcriber to represent the quantized local discourse structure with a sufficient level of confidence.

In various studies, quanta of local discourse structure are called syntagms, intonational phrases, intonational groups, rhythmic groups, intonation units, prosodic phrases/units/constituents, basic discourse units, among others; see, for example, Chafe (1994, p. 57, 2001); Cruttenden (1986); Degand & Simon (2009); Domínguez, Farrús, & Wanner (2016); Krivnova (2016); Stelma & Cameron (2007); Svetozarova, Vol'skaja, Pavlova, & Shitova (1988); Ščerba (1955); Xitina (2004); Yanko (2008). Quanta have been identified both manually and (semi)automatically on the basis of various criteria (prosodic, grammatical, semantic, etc.) and various approaches, including instrumental, experimental and corpus-based. Some frameworks propose "flat" subdivision of spoken discourse into minimal units, while others advocate a hierarchical approach allowing intermediate levels, see Shlomo Izre'el's consecutive segmentation into intonation units, paratones and periods (Izre'el & Mettouchi, 2015, and related work).

We use the term *elementary discourse unit* (EDU), thus emphasizing the constructional role of these units with respect to discourse production. Dividing discourse into EDUs is an important task, and such division is the most fundamental element of our transcription system. In a transcript, a graphic line corresponds to an EDU. Other important phenomena of local discourse structure represented in our transcription system include illocutionary types, phase, disfluencies and pausing. In our transcription system, we represent multiple prosodic phenomena, including primary (rhematic) and other accents, pitch direction in accents, emphasis, reduction, tempo variation, tonal registers, among others.

This paper is structured as follows. Following the editors' guidelines, we discuss the monologic and multi-party material somewhat separately (even though we argue against the strict monologue vs. dialogue distinction in our "Russian" paper; see Kibrik, Korotaev, & Podlesskaya, this volume, Part I). Sections 2 and 3 only address those phenomena that are not exclusively associated with multi-party exchange. In Section 2, we start off with the simplest ("canonical") instances and illustrate them with examples from both the monologic and the multi-party excerpts from the Santa Barbara corpus. In Section 3, we proceed with more complex phenomena, deviating from the canonical system, such as non-clausal EDUs and disfluencies. In Section 4, we proceed with those problems that only show up in multi-party discourse. Main conclusions are formulated inSection 5. Appendix A is a concise explanation of our transcription conventions, and Appendices B and C contain the transcripts of the two excerpts under analysis, adapted to the conventions we propose.

2. Basics

EDUs are identified primarily on prosodic grounds, including the following criteria: pausing pattern; holistic tonal contour; presence of an accentual center; loudness pattern; and tempo pattern. Figure 1 illustrates EDU Navy_E011 (see Appendix B), as represented by Praat (Boersma, 2001; Boersma & Weenink, 2012).

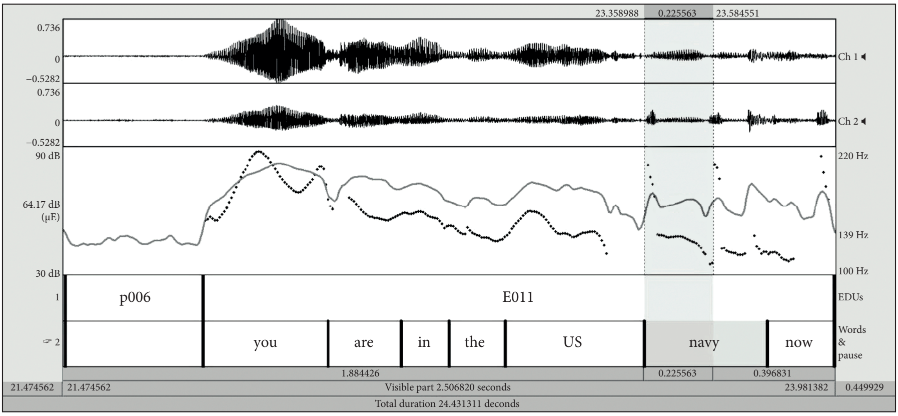


Figure 1. Acoustic representations of EDU Navy_E011

In Figure 1 one can see how the first four criteria work. First, there is a substantial pause before the beginning of the EDU Navy_E011; note that out of 11 EDUs in the *Navy* excerpt seven are preceded by an unfilled pause (see Section 4 on the interpretation of boundary pauses, as suggested by the data of multi-party exchange). Physiologically, boundary pauses are used for inhalation; cognitively, they allow a speaker to plan the upcoming EDU. Second, there is a holistic tonal contour in Navy_E011, see the speckles graph in Figure 1. (Note that artifacts of acoustic analysis may create a false impression that the contour is interrupted, such is the interruption during the pronunciation of [s] in *US*.) Third, there are clear pitch movements associated with accents, for example on the stressed syllable of *navy* (marked with a vertical stripe in Figure 1). Fourth, there is a gradual downdrift of intensity (the solid line graph in Figure 1); this effect is not as clear here as it should be because the listener's laughter overlaps with the speaker's talk during the last two words of the EDU. The final criterion, namely the tempo pattern, cannot be illustrated with Navy_E011, as the first word *You-u* is emphasized and lengthened (see Section 3), which is a strong confounding factor. Consider, however, the EDU Navy_E007. It is a complicated instance in itself (see Section 3), but it can be used to illustrate the point. The first two words of that EDU, that is *my friend*, and the last two words *his desk* (containing two syllables and seven phonemes each) have the durations of

540 ms and 640 ms, respectively, that is, there is a clear deceleration effect. See also Kibrik, Korotaev, & Podlesskaya (this volume, Part I, Section 3.1) for more details.

Apart from prosodic unity, EDUs display integrity at other levels, too. Cognitively, they represent a focus of consciousness, in terms of Chafe (1994). Semantically, they most typically convey an event (e.g., Navy_E007 or Hearts_J-E004) or a state (e.g., Navy_E011) and constitute nodes in a semantic network of discourse, for example, in terms of Rhetorical Structure Theory (Litvinenko, Podlesskaya, & Kibrik, 2009; Mann & Thompson, 1988). Grammatically, they tend to correlate with clauses. There are four clausal EDUs in the *Navy* excerpt and six clausal EDUs in the *Hearts* excerpt.

Certain words (or, rather, constituents) bear *accents*, that is, they are pronounced with more prominence. Accents are typically realized with pitch movements. For example, Navy_E002 contains two words with pitch accents: */thought* and *\friend*. Among the accents of an EDU, there is typically one that is functionally privileged, that is marks the most important ("rhematic") information offered in the EDU. We call this kind of an accent the *primary* one. For example, in Navy_E002 the word *\friend* bears the primary accent, and that makes sense as the speaker's communicative goal in this EDU is to challenge his friendship with the captain.

Direction of pitch in primary accents relates to the discourse-semantic category of *phase* (a term from Kodzasov, 1996, 2009). Phase is an abstract semantic category "anticipated continuation versus end" of something. Pitch accents convey this semantics iconically: A rising accent means "anticipated continuation, non-final", and a falling accent "end, final". Phase can be seen at least at three different hierarchical levels.

First, at the level of communicative exchange certain illocutions are final, that is, do not necessarily project a continuation – in particular, statements. Other illocutions anticipate a reaction and are thus non-final, for example, yes/no questions. Consider the EDUs Hearts_J-E001 and Hearts_D-E005; the first one is a yes/no question and the second one is a statement (that is, an answer to the question). Accordingly, we observe a rising accent in Hearts_J-E001 (*/hearts*) and a falling accent in Hearts_D-E005 (*\play*). Another clear instance of a statement is Navy_E011.

Second, there is a level of illocution-internal phase, close to the notion of transitional continuity in Du Bois, Schuetze-Coburn, Cumming, & Paolino (1992). EDUs form illocutionary chains that are spoken analogs of written sentences. (On operational approaches, allowing one to overcome obstacles associated with the notion of "sentence" as applied to spoken discourse, see, inter alia, Pietrandrea, Kahane, Lacheret, & Sabio, 2014.) Illocution-non-final EDUs tend to obtain a primary accent with the direction of pitch that is a mirror image of the pitch accent of the subsequent illocution-final EDU. In particular, statement-concluding EDUs are typically marked by falling primary accents, for example, Hearts_J-E004;

accordingly, non-statement-final EDUs usually have a rising primary accent, see Hearts_J-E003. In our transcription system, the illocutionary function of a chain is indicated with a closing punctuation mark on the illocution-final EDU; in particular, periods, question marks and inverted exclamation marks are used to signal statements, questions and directives, respectively. Illocution-non-final EDUs are, in a default case, closed with a comma. See Appendix A and Kibrik, Korotaev, & Podlesskaya (this volume, Part I) for more details.

Third, there is a level of EDU-internal phase. An EDU's primary accent typically appears towards the end of the EDU. If there is another preceding accent in the EDU, it usually adapts in the mirror-image way to the primary accent, see Navy_E002.

3. More complex instances

The basic system of local discourse structure and the corresponding transcription conventions, introduced in Section 2, only account for the most canonical instances. There are numerous complications to this basic system. There is no space here to address all of those, so only the most salient ones are briefly considered in this section.

Of course, it is not always the case that all five prosodic criteria of EDU identification converge. For example, there are no boundary pauses between Navy_E004 and Navy_E005, as well as between Navy_E008 and Navy_E009. Apparently, pairs of EDUs can be produced at one exhalation. The loudness downdrift pattern is not observed in Navy_E004, which may be partly accounted for by the introductory and subsidiary character of the quotative expression *I said* at the beginning of the EDU.

We differentiate between several types of EDUs from the point of view of their content. As has been shown above, canonical EDUs are clausal. *Subclausal* is an EDU that semantically belongs to an adjacent clausal EDU (so-called base clause), but prosodically is realized separately. Among the subclausal EDUs by far the most common are retrospective ones that follow the corresponding base clauses. An example is found in Navy_E008; this is an *increment*, that is an attribute/adjunct semantically belonging to the preceding base EDU. In the *Navy* excerpt there is also an instance of *split* – a structure in which a clause is divided into two subclausal units (Navy_E001 and Navy_E003); the first one contains the predicate *went to*, and the second contains its object *this captain*; the repeated demonstrative *this* helps to link the two parts of the split. Between the two parts there is an inset EDU, in this case Navy_E002 that is a relative clause.

Superclausal EDUs are formed when two or more predicative words are found in a single prosodic complex. The most common source of superclausal EDUs are complement constructions in which the syntactic matrix clause is prosodically tightly linked with the subordinate clause. In the two excerpts under analysis, there are several superclausal EDUs, including two epistemic complement constructions (Hearts_D-E005 and Navy_E002), a quotative construction (Navy_E004), and a modal verb construction (Navy_E006).

Paraclausal EDUs are those in which content is not propositional. There are several subtypes, including (but not limited to) holophrases (Hearts_D-E004, Hearts_J-E003), interjections (Hearts_J-E002), and vocatives (Navy_E010).

Division into EDUs by well-trained transcribers usually leads to a high degree of inter-transcriber agreement. However, there are equivocal instances that not only can be transcribed differently by different experts, but are inherently dubious. One instance of that is found in Navy_E003. Recall that it is the second part of a split. One can suggest an EDU boundary after the word *captain*, which is by all means possible syntactically. Under that interpretation, *down at the naval headquarters* would be a separate subclausal EDU (increment). However, the accent on *captain* is not quite strong enough for serving as a primary accent of an EDU. Therefore, we use the single EDU interpretation as the main one, and indicate a possible seam with a special symbol ¦. Another instance of an EDU that can be divided into two on an alternative analysis is found in Navy_E007.

There is a wide gamut of disfluencies that disturb speakers' "ideal delivery" (Clark & Clark, 1977). Basically, we differentiate between mild and severe disfluencies. Mild disfluencies allow a speaker to preserve an EDU's integrity, such as an EDU-internal silent hesitation pause in Navy_E006 (after *we*). They can also be realized as filled pauses (such as a sequence of uh- and um-pauses at the beginning of Navy_E003, marked (ə) and (ɯ), respectively), phoneme lengthening (*the-e* in Navy_E003), interrupted words (*m=* in Navy_E002, *n*^{ə}*= n=* in Navy_E003), or word repetitions (*and* || *and* in Navy_E006). Severe disfluencies, otherwise called *false starts* or *repairs*, take place when a speaker drops a constituent or a whole EDU. An example, although hard to hear, seems to appear in Hearts_J-E005; see Podlesskaya (2015) for a detailed account of disfluencies.

Apart from the canonical "comma intonation" with the rising primary accent, speakers also employ an alternative – a "falling comma intonation", or *non-final falling*. It can be identified as distinct from the "period intonation" (final falling). In the case of final falling, the intonation contour aims at the very bottom of an individual speaker's voice f0 range, while in non-final falling the target frequency is several semitones higher. In the excerpts under analysis the most clear example of final falling is found in the EDU Navy_E011; that gives us a hint of what is the bottom of the speaker's frequency range: about 100 Hz. In comparison to that, the *Navy* excerpt contains several examples of non-final falling (Navy_E002, E005, E007, and E008), in all of which the intonation contour targets the level of between 115 and 120 Hz, that is, 2.4 to 3.2 semitones higher than in Navy_E011.

Other peculiar tonal phenomena to mention include accents with complex pitch movements: /*could* in Navy_E006 or \/*hearts* in Hearts_D-E003. Also consider a highly marked contour on ↑ ↓/↑ *Lieutenant* in Navy_E010: There are salient pitch movements outside the stressed syllable – fall and rise on the pre-stressed syllable, and continued rise on the post-stressed syllable.

Some speakers occasionally use high tempo of speech. One instance is found in Hearts_D-E005 containing the high tempo sequence *I don't know how to*. That sequence involves five syllables and lasts for about 500 ms, so the mean syllable rate is 100 ms. Compare this with the normal tempo of this speaker, found for example in Hearts_D-E003, where the seven syllables sequence *I've never played \/\hearts before* lasts for about 1.2 s, which gives the mean syllable rate of 170 ms.

Finally, a salient feature of the speaker of the *Navy* excerpt is the frequent use of emphatic pronunciation, for example /\/***You-u*** in Navy_E011. Emphatic pronunciation is particularly articulate and loud and may be accompanied by lengthening of vowels.

4. Challenges of multi-party discourse

Transcribing multi-party discourse requires some additional conventions, also providing useful insights into how monologues should be transcribed. If our analysis were exclusively based on monologic discourse, boundary pauses could have been included in the upcoming EDUs. That would make sense for the reasons stated above, particularly because it is during this interval of time that a speaker prepares an upcoming EDU. However, in the case of multi-party discourse this principle does not work. There is no way to distinguish between a speaker's silence (his/her own boundary pause) and his/her silence in the role of a hearer. For example, consider the beginning of EDU Hearts_D-E006. There is a period of shared silence before it (0.37 s). But it would be wrong to interpret this pause exactly as Dan's boundary pause, as he was silent during a much longer time, since the end of his EDU Hearts_D-E005. So we have to recognize that in multi-party discourse there are no individual timelines. There is a shared timeline, and periods of shared silence must be treated as separate events, on a par with participants' EDUs. Of course, this does not concern EDU-internal pauses, as well as EDU-initial filled pauses (there are no examples in *Hearts*).

For multi-party discourse, it is convenient to organize a transcript in the form of a table with several columns, one for shared pauses and one per each participant's contributions, see the transcript of *Hearts* in Appendix C. This is a format sometimes called "scores" transcript. In fact, this representation should be seen as the canonical standard, while the representation we use in the transcript of *Navy* (where boundary pauses are marked on separate lines but in the same column) is a simplification that can only be applied to pure monologues.

The scores format is also useful for transcribing such a widespread phenomenon of multi-party discourse as *overlap*. A relatively simple instance of overlap is found between the EDUs Hearts_D-E003 and Hearts_J-E001, where Jenny's contribution starts 0.44 s before the end of Dan's EDU. Accordingly, there is no shared pause between these two EDUs. A more complex instance is found around EDU Hearts_J-E005. The initial part of this EDU overlaps with the end of the preceding Dan's EDU, and the final part with the whole of the subsequent Dan's EDU. In order to distinguish two different overlaps, we use slightly different symbols: a single bracket symbol versus double brackets. Note that in actual talk it is often difficult to precisely identify the boundaries of overlaps. For example, in Hearts_D-E006 we indicate the beginning of the overlap with Jenny's talk between [b] and [l] in *disabled*, but the boundary may actually take place one phoneme to the left or to the right.

5. Conclusion

We have thus submitted our Russian-based approach towards local discourse structure to a test, having applied it, in an exploratory way, to the English data. The main conclusion to draw from this trial is that the system does work. Probably this is no big surprise, as our approach is ultimately derived from the one that was originally developed for the Santa Barbara corpus (Du Bois, Schuetze-Coburn, Cumming, & Paolino, 1992). However, over the course of years we have made numerous changes, both triggered by empirical Russian data and of a more general conceptual nature. So the applicability of our approach to English is still informative.

Anyhow, the resources of our approach, including transcription conventions, sufficed for the analysis of the English sample. Apparently, the organization of discourse structure and prosody is comparable across the two languages. To be sure, some elements do differ, as in any domain of language. That concerns some fine principles of accent placement, selection of pitch direction in accents, and the use of specific intonation contours. The dataset is too small to make big generalizations and propose typological parameters, these are just impressionistic notes about apparent differences of English from what prosodic devices could be used in Russian.

As we pointed out at the beginning, in this brief account we only addressed the most salient and recurrent phenomena observed in the English excerpts. Some other elements of our analysis can be seen in the transcripts (Appendices B and C). All of the used notation conventions are explained in Appendix A.

Acknowledgements

Research underlying this study was supported by grant # 19-012-00626 from the Russian Foundation for Basic Research.

References

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *GLOT International*, 5(9/10), 341–345.

Boersma, P., & Weenink, D. (2012). *Praat: Doing phonetics by computer* (Version 5.3.04) [Computer software]. Retrieved from <http://www.praat.org/>.

Chafe, W. (1994). *Discourse, consciousness, and time*. Chicago, IL: University of Chicago Press.

Chafe, W. (2001). The analysis of discourse flow. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 673–687). Malden, MA: Blackwell.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language*. New York, NY: Harcourt Brace Jovanovich.

Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.

Degand, L., & Simon, A. C. (2009). Mapping prosody and syntax as discourse strategies: How basic discourse units vary across genres. In D. Barth-Weingarten, N. Dehé, & A. Wichmann (Eds.), *Where prosody meets pragmatics* (pp. 81–107). Bingley: Emerald. https://doi.org/10.1163/9789004253223_005

Domínguez, M., Farrús, M., & Wanner, L. (2016). An automatic prosody tagger for spontaneous speech. In *Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers* (pp. 377–386). Osaka.

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1992). Discourse transcription. *Santa Barbara Papers in Linguistics*, 4, 1–225.

Fedorova, O.V., & Kibrik, A.A. (Eds.) (2020). *The MCD handbook: A practical guide to annotating multichannel discourse*. Moscow: Institute of Linguistics RAS.

Izre'el, S., & Mettouchi, A. (2015). Representation of speech in CorpAfroAs: Transcriptional strategies and prosodic units. In A. Mettouchi, M. Vanhove, & D. Caubet (Eds.), *Corpus-based studies of lesser-described languages. The CorpAfroAs corpus of spoken AfroAsiatic languages* (pp. 13–41). Amsterdam: John Benjamins.

Kibrik, A. A. (2011). Cognitive discourse analysis: Local discourse structure. In M. Grygiel & L. A. Janda (Eds.), *Slavic linguistics in a cognitive framework* (pp. 273–304). Frankfurt: Peter Lang.

Kibrik, A. A., Korotaev, N. A., & Podlesskaya, V. I. (this volume). Russian spoken discourse: Local structure and prosody. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Kibrik, A. A., & Podlesskaya, V. I. (Eds.). (2009). *Rasskazy o snovidenijax: korpusnoe issledovanie ustnogo russkogo diskursa* [Night dream stories: A corpus study of spoken Russian discourse]. Moscow: Jazyki slavjanskix kul'tur.

Kodzasov, S. V. (1996). Kombinatornaja model' frazovoj prosodii [A combinatory model of phrasal prosody]. In T. M. Nikolaeva (Ed.), *Prosodičeskij stroj russkoj reči* [Prosodic structure of the Russian speech] (pp. 85–123). Moscow: IRJA RAN.

Kodzasov, S. V. (2009). *Issledovanija v oblasti russkoj prosodii* [Studies in the field of Russian prosody]. Moscow: Jazyki slavjanskix kul'tur.

Krivnova, O. F. (2016). Prosodičeskoe členenie zvučaščego teksta: tesktovaja lokalizacija dyxatel'nyx pauz [Prosodic phrasing in spoken text: Localization of breathing pauses]. *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialog"*, 15(22), 340–354. Retrieved from <http://www.dialog-21.ru/media/3404/krivnovaof.pdf>

Litvinenko, A. O., Podlesskaya, V. I., & Kibrik, A. A. (2009). Analiz rasskazov o snovidenijax s tochki zrenija ierarxičeskoj struktury diskursa. In A. A. Kibrik, & V. I. Podlesskaya (Eds.), *Rasskazy o snovidenijax: korpusnoe issledovanie ustnogo russkogo diskursa* [Night dream stories: A corpus study of spoken Russian discourse] (pp. 431–462). Moscow: Jazyki slavjanskix kul'tur.

Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8, 243–281. https://doi.org/10.1515/text.1.1988.8.3.243

Pietrandrea, P., Kahane, S., Lacheret, A., & Sabio, F. (2014). The notion of sentence and other discourse units in corpus annotation. In T. Raso, & H. Mello (Eds.), *Spoken corpora and linguistic studies* (pp. 331–364). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61.12pie

Podlesskaya, V. I. (2015). A corpus-based study of self-repairs in Russian spoken monologues. *Russian Linguistics*, 39 (1), 63–79. https://doi.org/10.1007/s11185-014-9142-1

Stelma, J. H., & Cameron, L. J. (2007). Intonation units in spoken interaction: Developing transcription skills. *Text and Talk*, 27(3), 361–393. https://doi.org/10.1515/TEXT.2007.015

Svetozarova, N. D., Vol'skaja, N. B., Pavlova, A. V., & Shitova, L. F. (1988). Prosodičeskaja organizacija russkoj spontannoj reči [Prosodic organization of spontaneous Russian speech]. In N. D. Svetozarova (Ed.), *Fonetika spontannoj reči* [Phonetics of spontaneous speech] (pp. 141–182). Leningrad: Izdatel'stvo Leningradskogo Universiteta.

Ščerba, L. V. (1955). *Fonetika francuzskogo jazyka* [French phonetics]. Moscow: Izdatel'stvo literatury na inostrannyx jazykax.

Xitina, M. V. (2004). *Delimitativnye priznaki ustno-rečevogo diskursa* [Delimitative features of spoken discourse]. Moscow: MGLU.

Yanko, T. (2008). *Intonacionnye strategii russkoj reči v tipologičeskom aspekte* [Intonational strategies in spoken Russian from a comparative perspective]. Moscow: Jazyki slavjanskix kul'tur.

Appendix A. Transcription conventions

We here only discuss those conventions that are used in the transcript in Appendices B and C. A more detailed discussion can be found in Kibrik, Korotaev, & Podlesskaya (this volume, Part I).

Elementary discourse units (EDUs) are represented as separate lines in transcripts. The ¦ symbol indicates those instances in which an EDU, on a possible interpretation, could be divided into more than one EDUs.

EDU boundaries are annotated in Praat, as well as the boundaries of words and pauses. In the multi-party transcript (*Hearts*), we indicate the starting and end times of EDUs in two separate columns. In the monologue transcript (*Navy*), we only indicate the starting time, as each EDU ends exactly at the same time as the subsequent EDU, or a silent pause, starts.

In the case of *overlaps* in multi-party discourse, the end time of an EDU is marked on a different graphic line, so that the end times and start times are sorted in a unified ascending order. See, for instance, the case of Hearts_D-E003 and Hearts_J-E001: The starting time of Dan's contribution is marked on line 0005; the starting time of Jenny's EDU is marked on line 0006; the end time of Dan's EDU is marked on line 0007; and the end time of Jenny's contribution is marked on line 0008. All these four time marks are aligned in an ascending order (2.12 – 3.33 – 3.77 – 4.47). The overlapping parts of EDUs are marked with [ ] and [[ ]] (the latter is used to avoid ambiguity when there is more than one overlap nearby).

Somewhat different principles of *pause* annotation are implemented for monologic and multi-party discourse, see Section 4 for discussion. In multi-party discourse, periods of shared silence are indicated in separate columns placed to the left of the shared timeline. Intervals of continuous vocalization by one speaker are shown with color filling.

In monologue transcripts, boundary pauses are marked on separate lines, but in the same column as EDUs. Contrary to silent pauses, filled pauses are always attributed to a particular speaker. For each pause, its duration is indicated in the transcript. The symbols used to represent pauses are indicated in Table 1.

Table 1. Symbols used to represent pauses in transcripts

(0.25*)	silent (no vocalization) pause
(ɥ 0.37)	silent pause filled by a loud inhalation sound (ingressive air flow)
(ə 0.78)	*uh*-like filled pause
(ɯ 1.01)	*um*-like filled pause

* Numbers in parentheses indicate the duration of a pause, in seconds.

Conventions used for *pitch* and *pitch accents* annotation are as follows. Symbols /, \, –, (as well as combinations thereof) indicate the direction of pitch on the stressed syllable of a word bearing a discourse accent. Arrows (↑, ↓, →) indicate significant pitch movements before and after the stressed syllable.

An EDU may contain more than one discourse accent. In most EDUs, one accent can be characterized as the *primary accent*, see Section 2. The primary accent is graphically distinguished from the other accents by underlining the vowel of the stressed sy̲llable. Most EDUs end with a *punctuation mark* that indicates an EDU's phase-related properties, including its illocutionary function (= exchange-level phase) and transitional continuity (= illocution-internal phase); see Tables 2 and 3, respectively.

Table 2. Some punctuation marks used to indicate EDUs' illocutionary functions (exchange-level phase)

.	statement
…	vague statement, inexhaustiveness
¡	directive
?	question
@	vocative

Table 3. Punctuation marks used to indicate EDUs' non-final transitional continuity (illocution-internal phase)

,	default incompleteness (may be combined with a mark of illocutionary function)
:	elucidation, used in cataphoric introductions and direct quotations
,,,	open list continuity, inexhaustiveness combined with incompleteness*

* See Kibrik, Korotaev, & Podlesskaya(this volume, Part I, Section 6.6) for more details on … and ,,, marks.

Illocutionary and transitional continuity properties can be combined with exclamation, in particular: ! (statement), @! (vocative). Internally-induced *repairs*, or *false starts,* are marked differently inside EDUs (||) and at the end of truncated EDUs (==). See Kibrik, Korotaev, & Podlesskaya (this volume, Part I, Section 7) for details on externally-induced repairs. Other transcription symbols are represented in Table 4.

Table 4. Other transcription symbols

bold	emphasis
a-a, r-r	lengthening
italics	accelerated tempo
grey	perceptible phonetic reduction
an=	truncated word
{sm 0.15}	non-words (such as smacking) and their duration
—	split (a clausal structure is split into two parts as another EDU wedges in)
“ ”	direct or semi-direct speech
<gonna>	uncertain
<UNCLEAR: 2>	unintelligible: number of syllables

In addition to the transcript as such, in Appendices B and C we also provide:

a. Information on *EDU type* (see Section 3);
b. Information on the properties of the primary pitch accent in the given EDU (see Sections 2 and 3);
c. Miscellaneous comments.

Appendix B. Transcript of *Navy*

Time	Line ID	Transcription	EDU type	Primary accent	Comments
0.00	E001	/So I went to this —	First part of the split	No primary pitch accent	
0.84	p001	(0.41)			
1.24	E002	m= ‖ what I /thought was my \friend,	Superclausal: epistemic construction; inset inside the split	Non-final falling	
2.75	E003	— (ə 0.19) (ɯ 0.19) nᵊ= n= ‖ this navy /\captain ¦ down at the-e naval /↓headquarters,	Second part of the split	Rising	“down at the naval headquarters” may be a separate subclausal EDU; but the accent on “captain” is not quite strong enough for being a primary accent of an EDU
5.74	N001	(ɥ 0.58)			
6.32	p002	(0.68)			
7.00	N002	{sm 0.09}			
7.09	E004	I said “This is terribly /↓awkward,	Superclausal: quotative construction	Rising	
8.35	E005	/I’ve just been promoted from /**third** mate (0.13) to \second mate,	Clausal	Non-final falling	tapping on the desk

Time	Line ID	Transcription	EDU type	Primary accent	Comments
11.01	N003	(ɥ 0.50)			
11.52	E006	and || and /\could we (0.23) /\possibly (0.08) postpone these \orders for a little bit?".	Superclausal: modal verb construction	Final falling	tapping on the desk; semidirect speech: no imitation of interactional prosody
14.76	p003	(1.01)			
15.77	E007	/**My** friend stood \up | (0.1) ↓/**behind his \desk,**	Clausal	Non-final falling	"behind his desk" may be a separate subclausal EDU; but the part ending with "stood up" seems prosodically incomplete
17.91	p004	(0.15)			
18.06	E008	in his /\fu-ull \f-four \–stripes,	Subclausal: increment	Non-final falling	
19.95	E009	and \said:	Clausal	Non-final falling	
20.36	p005	(0.46)			
20.83	E010	"↑↓/↑**Lieutenant**@!	Paraclausal: vocative	Extra high rising	very loud
21.46	p006	(0.45)			
21.93	E011	/\\/**You-u** are in the US \navy now^{h}."	Clausal	Final falling	"navy now" cooccurs with the listener's laughter

Appendix C. Transcript of *Hearts*

Line #	Pauses		TimeS	TimeE	Dan		Jenny		EDU type	Primary accent	Comments
0001	p001	(1.07)	0.00	1.07							
0002			1.07	1.36	D-E001	/\Wait¡,			Clausal	Non-final falling	
0003	p002	(0.13)	1.36	1.49							
0004			1.49	2.12	D-E002	play /\no-ovice¡,			Clausal	Non-final falling	
0005			2.12		D-E003	I've never played \/\hearts before <[in my life]>.			Clausal	Final falling	
0006			3.33				J-E001	[You've never] played /hearts?	Clausal	Rising	
0007				3.77							
0008				4.47							
0009	p003	(0.09)	4.47	4.58							
0010			4.56	4.82	D-E004	\No,			Paraclausal: holophrase	Non-final falling	
0011			4.82	5.68	D-E005	*I don't know how to* \play it.			Superclausal: epistemic construction	Final falling	

Line #	Pauses		TimeS	TimeE	Dan	Jenny	EDU type	Primary accent	Comments
0012	p004	(0.23)	5.68	5.91					
0013			5.91	6.37		J-E002 –\Oh!	Paraclausal: interjection	Final falling	
0014	p005	(0.31)	6.37	6.68					
0015			6.68	7.01		J-E003 ↓/Okay,	Paraclausal: holophrase	Rising	creaky voice
0016	p-006	(0.08)	7.01	7.09					
0017			7.09	7.62		J-E004 I’ll \teach you.	Clausal	Final falling	creaky voice
0018	p-007	(0.37)	7.62	7.99					
0019			7.99		D-E006 /Passing /\disab[led,]		Subclausal: no finite verb	Non-final falling	laughter on “passing”
0020			8.98			J-E005 <[Que][[en of sp^{h}= ==]]>	False start		
0021				9.14					
0022			9.14		D-E007 [[that’s \you.]]		Clausal	Final falling	
0023				9.62					
0024				9.66					

CHAPTER 7

Some notes on the *Hearts* and *Navy* excerpts according to the Language into Act Theory

Emanuela Cresti and Massimo Moneglia
University of Florence – LABLITA

The paper sketches the Language into Act Theory and how it catches the difference between the *Navy* monologue and the *Hearts* dialogue. According to L-AcT, two types of reference units, both ending with a prosodic terminal break are identified: *utterance* matching with a single speech act and *stanza* expressing a flow of thought through an adjunction process. *Navy* is a sequence of two narrative *stanzas* with a complex informational organization, while *Hearts* is organized in 11 *utterances* showing high illocutionary variation. The core of the information pattern is the Comment accomplishing the illocutionary force. The information structure, expressing a closed set of functions, is in one-to-one correspondence with the prosodic structure. The linguistic content is not compositional across information units.

Keywords: reference units, illocutionary force, information structure, prosodic structure, compositionality

1. Premises

The Language into Act Theory (L-AcT; Cresti, 2000) addresses the problem of identifying *speech reference units* in the linguistic analysis of speech. The primary unit of reference is the *utterance*, which is pragmatic in nature and the counterpart to a speech act, in keeping with the definition given by Austin (1962). L-AcT's main innovation is in how it considers the utterance to be necessarily performed and identifiable through *prosodic means*, while also corresponding to an *information pattern* which may be composed of many units displaying different information functions. The centre of the information pattern is constituted by a specific information unit known as the *Comment* (COM), which is dedicated to the accomplishment of the utterance's *illocutionary force* and is necessary and sufficient for performing an utterance.

https://doi.org/10.1075/scl.94.16cre

On the basis of the identification of the Comment unit, empirical research carried out on corpora allowed the collection of a repertory of illocutionary types, structured into five main classes, many sub-classes, and dozens of illocutionary types (roughly 90), which appear to be shared across English and Romance languages (Cresti, 2017). This repertory, presented in detail in Cresti (this volume), is a working set and open to the addition of new entries which may be discovered in the course of further corpus-based investigations.

Corpus-driven research has also led to the discovery of a second reference unit known as the *stanza* (Cresti, 2009). The stanza, too, is pragmatic in nature as it is constituted by a sequence of bound Comments (COB) each expressing some illocutionary value. A stanza does not correspond to a sum of utterances but is a different type of reference unit in which each Bound Comment may in turn be supported by other information units (information sub-patterns). Furthermore, contrasting with the information patterns for utterances, a sequence of COBs is not produced by the speaker as a whole, but is conceived "on the fly" and continues on until the end of the flow of thought.

The stanza somewhat approximates the conception proposed by Chafe (1970). Its pragmatic value is rather low since the illocutions of the COBs are mostly assertive and weak (i.e., characterized by weaker commitment to the truth of the contents and a low degree of involvement with the addressee). It is not by chance that stanzas occur especially in monologic and formal texts, examples of which may be seen in the second excerpt from our analysis.

According to previous studies in the L-AcT framework, the incidence of utterances and stanzas in spoken performance is quite different. For instance, Italian IPIC data records the utterance as approximately 90% and the stanza as 10% of total terminated prosodic sequences that formally correspond to reference units (Panunzi & Gregori, 2011).

Prosody delimits the boundaries of reference entities (both utterances and stanzas) through terminal breaks and is the necessary interface between the illocutionary act and the locutionary act. Beyond the utterance and the stanza, prosody also brings another typical spoken strategy to light: the *illocutionary pattern* (Panunzi & Saccone, 2018). Illocutionary patterns are chains of two or more Comments within a single utterance and do not constitute an additional reference unit type. They are conceived as a whole according to a natural rhetoric model (*reinforcement, list, alternative question, comparison, adversative proposal*) and produce a chain of rhythmed multiple Comments (CMM). The Comment can double up or occur three times, repeating the same illocutionary type or varying the illocutionary types in the same pattern.

As noted, an *information pattern* has its centre in the Comment information unit. If composed of this unit alone, the pattern is classified as *simple*. The Comment may also co-occur with other, additional information units and in such cases the pattern is said *complex*. Each information unit maps to a chunk of speech which is delimited by terminal or non-terminal prosodic breaks and is distributed in the utterance with respect to the Comment. Prosodic breaks are perceived in speech in correspondence to f0 reset, lengthening, pauses and drop of intensity (Izre'el, 2005; Moneglia & Cresti, 2006). Prosodic units are characterized by perceptively relevant movements of different types (Firenzuoli, 2003; 't Hart, Collier, & Cohen, 1990).

Information units are divided into two types: *textual* and *dialogic*. Textual units implement the semantic content of the utterance and must correspond to an identifiable semantic entity. The dialogic units are only devoted to the management of the communication itself and do not participate to the semantics of the utterance (Cresti, 2000; Moneglia & Raso, 2014; Raso, 2014). See Table 1 for the tag set of Information functions. According to L-AcT, the locutive content of each textual information unit constitutes a syntactic and semantic island with its own modality; that is, information units are not compositional at the syntactic and semantic levels (Cresti, 2014, 2019). Compositionality holds within the information units.

A chunk of speech between two prosodic breaks may not develop an information function in cases where the prosody simply divides into parts – that is, scans – an information unit that is too long to be performed as one prosodic unit. These *Scanning units* (SCA), fall into the list of textual information units, since only textual units such as the Comment, Topic, Parenthesis, and (rarely) the Appendix may be scanned. When scanning occurs, the fulfilment of one information function may be delayed (*scanning on the left*), or, less frequently, prolonged (*scanning on the right*). In the Romance languages, scanning is almost always on the left, that is, only the last part of the scanned unit conveys the information function in question. The semantic/syntactic content of scanning units is compositional.

Table 1. Tag set of information unit types

Type of unit	Name	Tag	Definition
Textual	Comment	COM	Accomplishes the illocutionary force of the utterance.
	Topic	TOP	Identifies the domain of application for the illocutionary act expressed by the Comment.
	Appendix of Comment	APC	Integrates the text of the Comment and concludes the utterance, indicating agreement with the addressee.

(*continued*)

Table 1. (*continued*)

Type of unit	Name	Tag	Definition
	Appendix of Topic	APT	Yields a delayed integration of the information given in the Topic.
	Parenthesis	PAR	Inserts information into the utterance with a meta-linguistic value.
	Locutive Introducer	INT	Expresses the evidence status of the subsequent locutive space, marking a shift in the coordinates for its interpretation.
	Multiple Comment	CMM	Constitutes a chain of Comments which form an *illocutionary pattern*, i.e., an action model which allows the linking of at least two illocutionary acts, for the performance of one conventional rhetoric effect.
	Bound Comment	COB	A sequence of Comments, which are produced by progressive adjunctions following the flow of thought (*Stanza*).
	Scanning Unit	SCA	Scans the locutive content of a textual information unit
Dialogic	Incipit	INP	Opens the communicative channel, bearing a contrastive value and initiating a dialogic turn or an utterance.
	Conative	CNT	Pushes the listener to take part in the Dialogue or to stop his uncollaborative behavior.
	Phatic	PHA	Controls the communicative channel and maintains it. Stimulates the listener toward social cohesion.
	Allocutive	ALL	Specifies to whom the message is directed while holding their attention and forming a cohesive, empathic function.
	Expressive	EXP	Functions as an emotional support, stressing the sharing of a social affiliation.
	Discourse Connector	DCT	Connects different parts of the discourse, indicating their continuation to the addressee.

2. The tagged transcription according to L-AcT

The following transcript is provided with tags for the information function of each of the information units and, below each reference unit, their illocutionary value. The reference units are utterances by default. Stanzas are identified by the presence of COB units and Illocutionary patterns by CMM units. Appendices A to F provide detailed analyses of each reference unit and of each information unit at the prosodic, pragmatic, informational, syntactic and semantic levels. Prosodic features reported in tables (Appendices A and D) are not explicitly discussed in the paper.

(1) *Hearts*

*DAN: *wait //*$^{\text{COM}}$ *play novice //*$^{\text{COM}}$ *I've never played Hearts before /*$^{\text{COM}}$ *<in my life> //*$^{\text{APC}}$
%ill: move (assent); move (waiting request); assertion (self-conclusion); assertion (admission)

*JEN: *<you've never> played hearts //*$^{\text{COM}}$
%ill: acknowledgment (with tired or sufficiency attitude)

*DAN: *no /*$^{\text{CMM}}$ *I don't know how to play it //*$^{\text{CMM}}$
%ill: explanation (reinforcement pattern)

*JEN: *oh //*$^{\text{COM}}$ *okay /*$^{\text{CMM}}$ *I'll teach you //*$^{\text{CMM}}$
%ill: acknowledgement; conclusion (reinforcement pattern)

*DAN: *passing disabled //*$^{\text{COM}}$ *<that's you> //*$^{\text{COM}}$
%ill: citation (reading); direction (passing the turn)

*JEN: *queen of &sp +*$^{\text{EMP}}$
%ill: 0

(2) *Navy*

* TOC: *so I went to this /*$^{\text{i-COB}}$ *what I thought was my friend /*$^{\text{PAR}}$ *this navy captain down at /*$^{\text{SCA}}$ *naval headquarters /*$^{\text{COB}}$
%ill: narration
I said /$^{\text{INT}}$ *this is terribly awkward //*$^{\text{COM-r}}$ *I've just been promoted /*$^{\text{COB-r}}$ *from /*$^{\text{INT-r}}$ *third mate /*$^{\text{CMM-r}}$ *to second mate /*$^{\text{CMM-r}}$ *and /*$^{\text{AUX-r}}$ *could we /*$^{\text{i-COM-r}}$ *possibly /*$^{\text{PAR}}$ *postpone these orders /*$^{\text{COM-r}}$ *for a little bit //*$^{\text{APC}}$
%ill: reported speech
my friend/$^{\text{TOP}}$ *stood up /*$^{\text{SCA}}$ *behind his desk /*$^{\text{COB}}$ *in his full /*$^{\text{SCA}}$ *&f four stripes /*$^{\text{COB}}$
%ill: narration
and said /$^{\text{INT}}$ *Lieutenant //*$^{\text{COM-r}}$ *you /*$^{\text{TOP-r}}$ *are in the US Navy now //*$^{\text{COM-r}}$
%ill: reported speech

Despite their brevity, these excerpts may be considered representative examples of the two main types of spontaneous spoken interaction: continuous exchanges between speakers (dialogue or multi-dialogue) and single-speaker acts (monologue).

3. The pragmatic analysis

Dialogues and monologues can be distinguished on the basis of pragmatic characters. For Examples (1) and (2), their difference may be demonstrated quite easily: The dialogue is composed of six dialogic turns occurring between two speakers but corresponds to the accomplishment of *11 utterances* + 1 interrupted unit (frequent in spontaneous speech) and shows a significant *pragmatic variation*, while the monologue is constituted of a single speaker's turn and is composed of two *narrative stanzas*. This datum confirms the crosslinguistic trend recorded in the IPIC data base: While in Italian monologues stanzas represent around 23% of reference units and in Brazilian monologues around 25%, stanzas are respectively 7.2% and 5.4% in dialogues (Panunzi & Mittmann, 2014).

In (1), the turns of speakers DAN and JEN may be appreciated through the change in voice (male vs. female), but it must also be noted that within each turn different utterances are accomplished which can be identified through their systematic correlation with terminal prosodic breaks. However, of more relevance is the continuous change in illocutionary types performed during the exchange by the two speakers. The 11 *utterances*, indeed, each correspond to a different illocutionary type.

The subject of the dialogue is the explanation of the card game "Hearts", being played on a computer. The first turn by DAN, who admits that he does not know how to play the game, is a sequence of different illocutionary acts (*assent, waiting request, self-conclusion, explanation, admission*) which follow his reactions and change in attitude toward.

The aforementioned illocutionary values arise out of the L-AcT repertory, that are more varied than Searle's traditional taxonomy (Searle, 1969). For instance, the first illocutionary type (*assent*), corresponds to a dialogical move with almost no semantic content, functioning to signal to the addressee that the speaker has understood what has been said to him (*neat* //$^{\mathrm{COM}}$). The subsequent *waiting request* is also a dialogical move that asks the addressee to stop speaking for a little bit (*wait* //$^{\mathrm{COM}}$). *Self-conclusion* is an assertive illocutionary type in which the speaker appears to talk to himself, justifying his prior statements (*play novice* //$^{\mathrm{COM}}$) (see Cresti, this volume, Part I). Lastly, the *explanation* and *admission* types clarify the speaker's ability for the addressee (*I've never played Hearts before* /$^{\mathrm{COM}}$ <*in my life*> //$^{\mathrm{APC}}$).

It is worth noting the reaction of the speaker JEN, who intends to play the game with him and discovers that he is unable to, which she *acknowledges* with an attitude of slight impatience and boredom (<*you've never*> *played hearts* //$^{\mathrm{COM}}$). DAN then reaffirms his previous admission, explaining that he is unable to play the game with a typical illocutionary pattern of *reinforcement* (*no* /$^{\mathrm{CMM}}$ *I don't know how to play it* //$^{\mathrm{CMM}}$).

This time JEN fully acknowledges the fact (oh //$^{\mathrm{COM}}$) (see Cresti, this volume, Part I), and concludes by encouraging a proposal to teach, behaving too with an illocutionary pattern of *reinforcement* composed of two *conclusions* (*okay* /$^{\text{CMM}}$ *I'll teach you* //$^{\text{CMM}}$). Reinforcement, indeed, is the most common form of illocutionary pattern and is achieved by repeating the same illocutionary type. Other common illocutionary patterns are *compariso*n and *alternation*, as well as chains of three or more Comments, which constitute a *list*.

Now that the game can start, DAN reads aloud the instructions displayed on the screen, accomplishing a *citation* illocution (*passing disabled* //$^{\text{COM}}$). Then, it is JEN's turn and DAN gives his *agreement*, pushing her to play (<*that's you*> //$^{\text{COM}}$). Finally, JEN says something (probably reading her card from the screen), but does not end her speech, leaving it fragmented.

The performance of each illocutionary act in the *Hearts* example corresponds to a reference unit of the *utterance* type and is mostly characterized by a dedicated prosodic profile. No *stanza* is performed. The linguistic content of (1) is limited to few words while the dialogue is realized through subtle and complex psychological dynamics between the speakers, giving rise to a continuous variation in illocutionary acts, most of which cannot be predicted from the contextual information.

The *Navy* excerpt in (2) presents a very different kind of pragmatic exchange. Even though the speaker TOC is at dinner with good friends, who ask questions and make comments, he is still the "dominant" speaker. Example (2) corresponds to a singular turn example, in which TOC tells of when and how he entered the US Navy. The tale is unitary but composed of two stanzas which are both structured in the same way: a *narrative introduction* followed by a stretch of *reported speech*, containing some alleged mimetic reproductions of other speakers. In turn, both the introductions and the reported utterances are composed of different occurrences of illocutionary weak assertive and directive types (see Cresti, this volume, Part I).

TOC demonstrates he is an expert in storytelling; long descriptions can lose the audience's attention and reported speech is a common device for re-energising a story. The act of reporting performed by TOC is not considered part of the assertive illocutionary class, but rather is assigned to the directive class since this form of "theatrical" performance or dramatization should not be judged in terms of its truth-value. The reported speech act enacts a kind of request, aimed at modifying the addressee's mind and pushing him to accept what is reported as reality. Taken as a whole, (2) corresponds to a sequence of two *stanzas*, which are composed both of episodes of assertive narration and of reported speech.

While the content is the relevant thing in a monologue, the dynamics between the speakers, though less direct than in a dialogue, are still there. The speaker must look at his addressees and even if silent, must take them into consideration. Therefore, the monologue's pragmatic characterisation is mostly reduced to assertive types, although these may be interspersed with expressive illocutionary acts and weak requests (as reported speech must be considered), encouraging the addressees to participate in the display of narration.

In conclusion, the pragmatic behaviour enacted in dialogues and monologues are basically different, testing different speaker abilities in performing speech.

4. The organization of information

The second basic aspect for distinguishing dialogues and monologues concerns the way in which they develop information structure. The utterances in (1) correspond in most cases to a *simple information pattern*, composed of a simple Comment accomplishing an illocution (*neat //*$^{\text{COM}}$ *wait //*$^{\text{COM}}$ *play novice //*$^{\text{COM}}$). Only the fourth utterance may correspond to a *complex information pattern*, being composed of two information units (*Comment-Appendix*) in which the first unit accomplishes an explanation force and the second, in our preferred interpretation, provides additional, essentially irrelevant information (*I've never played Hearts before /*$^{\text{COM}}$ <*in my life*> //$^{\text{APC}}$). The Comment couples of the sixth and seventh utterances and the eighth and ninth utterances are illocutionary patterns made up of two simple Comments, reinforcing the same illocution: *explanation* and *conclusion,* respectively (*no /*$^{\text{CMM}}$ *I don't know how to play it //*$^{\text{CMM}}$; *okay /*$^{\text{CMM}}$ *I'll teach you //*$^{\text{CMM}}$). In summary, the *information structure* of (1) is extremely simple and, within the interactive dynamics, functions to accomplish the illocution or to reinforce it.

Conversely, the *Navy* example presents quite a complex information organization. Before proceeding, we note that it is not by chance that all examples of scanning presented are taken from the monologue, for example, scanning on the left: *so I went to this /*$^{\text{i-COM}}$ *what I thought was my friend /*$^{\text{PAR}}$ *this navy captain down at /*$^{\text{SCA}}$ *naval headquarters /*$^{\text{COB}}$*; in his full /*$^{\text{SCA}}$ *&f four stripes /*$^{\text{COB}}$. English challenges the exclusive occurrence of this kind of scanning, since it presents cases of right scanning. In our opinion, a kind of collapse between Appendix units and right scanning may be observed. Indeed, Appendix of Comment units may be modelled in a way that could also be considered forms of right scanning. For instance, the uncertainty of the non-terminal break and weak semantic relevance in the following examples *I've never played Hearts before /*$^{\text{COM}}$ *<in my life> //*$^{\text{APC}}$ and *and /*$^{\text{AUX-r}}$ *could we /*$^{\text{i-COM-r}}$ *possibly /*$^{\text{PAR}}$ *postpone these orders /*$^{\text{COM-r}}$ *for a little bit //*$^{\text{APC}}$ allow their interpretation as right scanning. Nonetheless, we choose the Appendix tag since the low f0 profile and weak intensity are strictly consistent with the prosodic performance of the Appendix function.

As a whole, (2) corresponds to two stanzas, each of which records a rather complex information pattern and follows the same schema. The first stanza contains two parts:

1. The first Bound Comment is composed of a Comment interrupted by a Parenthesis and the Comment's continuation scanned in two pieces: *so I went to this /*$^{\text{i-COM}}$ *what I thought was my friend /*$^{\text{PAR}}$ *this navy captain down at /*$^{\text{SCA}}$ *naval headquarters /*$^{\text{COB}}$. The Bound Comment accomplishes a *narration* illocution;
2. The second part corresponds to an *episode of reported speech* and contains a Locutive introducer which introduces the mimetic reproduction of a stretch of speech, the information tags of which are super-scribed with "-r": *I said /*$^{\text{INT}}$ *this is terribly awkward /*$^{\text{COB-r}}$. After the first reported Bound Comment the reported speech continues on with two other reported Bound Comments, the first of which has a reported illocutionary pattern of *comparison* (*I've just been promoted /*$^{\text{COB-r}}$ *from third mate /*$^{\text{CMM-r}}$ *to second mate /*$^{\text{CMM-r}}$). It ends with the third reported Bound Comment, the information pattern of which contains a Parenthesis followed by an Appendix (*and /*$^{\text{AUX-r}}$ *could we /*$^{\text{i-COM-r}}$ *possibly /*$^{\text{PAR}}$ *postpone these orders /*$^{\text{COM-r}}$ *for a little bit //*$^{\text{APC}}$). All the Bound Comments are in a paratactic relation with one another.

The second stanza repeats the previous schema:

1. The first part is constituted by two Bound Comments. The first corresponds to an information pattern, which is composed of a Topic and a scanned Comment. This is in a paratactic relation with a second scanned COB (*my friend/*$^{\text{TOP}}$ *stood up /*$^{\text{SCA}}$ *behind his desk /*$^{\text{COB}}$ *in his full /*$^{\text{SCA}}$ *&f four stripes /*$^{\text{COB}}$). The Bound Comments accomplish a *narration* illocution.
2. In this case, too, the stanza is concluded by an *episode of reported speech* that is composed of two reported Comments. The first one corresponds to an information pattern composed of a Locutive Introducer and a reported Comment (*and said /*$^{\text{INT}}$ *Lieutenant //*$^{\text{COM-r}}$), and the second of a reported utterance, corresponding to a Topic-Comment pattern (*you /*$^{\text{TOP-r}}$ *are in the ues Navy now //*$^{\text{COM-re}}$). The whole episode accomplishes an illocution of *reporting*.

The stanzas of (2) continue to add pieces to the story. Beyond the apparent complexity of the information structure, the development of the tale is fluid, notwithstanding one phonetic misstep and some slight stuttering. The speaker clearly enjoys recounting the story and knows how to draw it out, thus he is behaving naturally and with little effort while making the tale fully understandable and accessible.

In summary, the basic pragmatic differences between the dialogue and the monologue are overtly reflected in their information structure, thus giving rise: in the first, to a sequence of simple utterances, composed of only a Comment or an illocutionary reinforcement pattern but showing strong illocutionary variation; and corresponding in the second to a single, long turn, composed of two stanzas, with significant internal information composition but low illocutionary variation.

The reference units and information pattern types in the dialogue and monologue differ; however, one should not be misled by this since complex instances of information patterning may appear frequently in spontaneous dialogue.

5. L-AcT analysis beyond pragmatics

The previous analysis is limited to some basic aspects investigated by the LABLITA team for spoken texts: pragmatics, reference units and information structure. All of these are identified via their prosodic profile and demarcation. However, Appendices C and F show other important levels that have also be considered, in order to analyse the two spoken texts in an integrated and organic way, that is the semantic and modal characterization of information units, and their syntactic fulfilment.

L-AcT assumes that the locutive content of each information unit matches with an independent semantic/syntactic island (Cresti, 2014). Within the utterance, islands are bound to each other in accordance with combination principles and do not give rise to compositional syntactic configurations. In the present examples, this assumption may be appreciated in (2), which presents a longer and more complex text than (1). For instance, in the first stanza: *I said /*$^{\text{INT}}$ *this is terribly awkward //*$^{\text{COM-r}}$ *I've just been promoted /*$^{\text{COB-r}}$ *from /*$^{\text{INT-r}}$ *third mate /*$^{\text{CMM-r}}$ *to second mate /*$^{\text{CMM-r}}$ *and /*$^{\text{AUX-r}}$ *could we /*$^{\text{i-COM-r}}$ *possibly /*$^{\text{PAR}}$ *postpone these orders /*$^{\text{COM-r}}$ *for a little bit //*$^{\text{APC}}$. This chunk corresponds to the introduction of a reported episode which is made up of a set of semantic/syntactic islands, which are neither completive subordinate clauses of the introducing VP (*I said /*$^{\text{INT}}$) nor coordinate clauses with each other. The information units, demarcated by prosodic breaks, are characterized by specific profiles that mimetically reproduce different illocutionary forces (the expression of evaluation, a description, a comparison pattern, a kind request), even while participating in the accomplishment of an overall reporting illocution. Each information unit is conceived in an autonomous way and is added and bound to the others in a *paratactic* manner (Cresti, 2019).

References

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

Chafe, W. (1970). *Meaning and the structure of language*. Chicago, IL: University of Chicago Press.

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

Cresti, E. (2014). Syntactic properties of spontaneous speech in the Language into Act Theory: Data on Italian complements and relative clauses. In T. Raso & H. Mello (Eds.), *Spoken corpora and linguistics studies* (pp. 365–410). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61.13cre

Cresti, E. (2017). The empirical foundation of illocutionary classification. In A. De Meo & F. Dovetto (Eds.), *Atti SLI – GSCP international conference, la comunicazione parlata* (pp. 243–264). Napoli: Aracne.

Cresti, E. (this volume). The pragmatic analysis of speech and its illocutionary classification according to Language into Act Theory. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Cresti, E. (2019). Aspetti interpuntivi nella prosa letteraria. In A. Ferrari, L. Lala, F. Pecorari, & R. Stojmenova Weber (Eds.), *Atti del convegno internazionale "Punteggiatura, sintassi, testualità nella varietà dei testi contemporanei"* (pp. 349–362). Firenze: Cesati.

Firenzuoli, V. (2003). Le forme intonative di valore illocutivo dell'Italiano parlato: Analisi sperimentale di un corpus di parlato spontaneo (LABLITA) (Unpublished doctoral dissertation). Università di Firenze, Italy.

't Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual study on intonation. An experimental approach to speech melody*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511627743

Izre'el, S. (2005). Intonation units and the structure of spontaneous spoken language: A view from Hebrew. In C. Auran, R. Bernard, C. Chanet, A. Colass, A. Di Christo, C. Portes, A. Reynier, & M. Vion (Eds.), *Proceedings of the IDP05 international symposium on discourse-prosody interfaces*. Aix-en-Provence: Université de Provence.

Moneglia, M., & Cresti, E. (2006). C-ORAL-ROM prosodic boundaries for spontaneous speech analysis. In Y. Kawaguchi, S. Zaima, & T. Takagaki (Eds.), *Spoken language corpus and linguistics informatics* (pp. 89–114). Amsterdam: John Benjamins. https://doi.org/10.1075/ubli.5.07mon

Moneglia, M., & Raso, T. (2014). Notes on the Language into Act Theory. In T. Raso & H. Mello (Eds.), *Spoken corpora and linguistics studies* (pp. 468–494). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61.15mon

Panunzi, A., & Gregori, L. (2011). DB-IPIC. AN XML database for the representation of information structure in spoken language. In H. Mello, A. Panunzi, & T. Raso (Eds.), *Pragmatics and prosody. illocution, modality, attitude, information patterning and speech annotation* (pp. 133–150). Firenze: Firenze University Press.

Panunzi, A., & Mittmann, M. (2014). The IPIC resource and a cross-linguistic analysis of information structure in Italian and Brazilian Portuguese. In T. Raso & H. Mello (Eds.), *Spoken corpora and linguistic studies* (pp. 129–151). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61.05pan

Panunzi, A., & Saccone, V. (2018). Complex illocutive units in L-AcT: An analysis of non-terminal prosodic breaks of bound and multiple comments. *Revista de Estudos da Linguagem*, 26 (4), 1647–1674.

Raso, T. (2014). Prosodic constraints for discourse markers. In T. Raso & H. Mello (Eds.), *Spoken corpora and linguistic studies* (pp. 411–467). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61.14ras

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139173438

Appendix A.

Table 1. Annotation of the *Navy* excerpt. Prosody

NAVY (1)			**LEVEL A: prosody**							
			A1 prosodic break	**A2 break features**				**A3 prosodic unit features**		
SPEAKER	**TURN**	**SEGMENT**	**Break type**	**A2.1 reset**	**A2.2 lengthning**	**A2.3 pause**	**A2.4 int-drop**	**A3.1 unit-type**	**A3.2 unit-structure**	**A.3.3 Perceptively relevant F0-mov**
TOC	1	so I went to this &m	/	p	n	n	p	i-root	0	0
TOC	1	what I thought was my friend	/	n	p	n	n	parent	unstructured	Flat
TOC	1	this navy captain down &dn	/	p	n	n	p	scan-root	0	0
TOC	1	at naval headquarters	/	p	n	p	p	b-root	prep-nucl	r/f
TOC	1	I said	/	p	n	n	n	intro	unstructured	(s)f
TOC	1	this is terrybly akward	//	p	n	n	n	root	prep-nucl	r/f
TOC	1	I've just been promoted	/	p	n	n	p	b-root	nucl-tail	r
TOC	1	from third mate	/	p	n	n	p	chained-root	nucl	r/f
TOC	1	to second mate	/	p	n	p	p	chained-root	nucl	f
TOC	1	and [/1] and	/	p	n	p	p	connective	unstructured	Flat

Table 2. Annotation of the *Navy* excerpt. Prosody (2)

NAVY			LEVEL A: prosody							
			A1 prosodic break	A2 break features				A3 prosodic unit features		
SPEAKER	TURN	SEGMENT	Break type	A2.1 reset	A2.2 lengthning	A2.3 pause	A2.4 int-drop	A3.1 unit-type	A3.2 unit-structure	A.3.3 Perceptively relevant F0-mov
TOC	1	could we	/	p	n	n	p	i-root	nucleus	f
TOC	1	possibly	/	p	n	n	p	parent	unstructured	Flat
TOC	1	postpone these orders	/	n	n	n	n	root	tail	Flat
TOC	1	for a little bit	//	p	n	p	p	suFFix	unstructured	Flat
TOC	1	my friend	/	p	n	n	n	preFix	nucleus	r/f
TOC	1	stood up	/	p	n	n	p	b-root	nucleus	f
TOC	1	behind his desk	/	p	n	p	p	scan-root	0	0
TOC	1	in his full	/	p/n	n	p	p	b-root	prep-nucl	f
TOC	1	&f four stripes	/	P/n	n	n	p	b-root	0	0
TOC	1	and said	/	p	n	p	p	intro	unstructured	f
TOC	1	Lieutenant	//	p	n	p	p	root	nucleus	r/f
TOC	1	you	/	p	p	n	p	preFix	nucleus	r
TOC	1	are in the US Navy now	//	p	n	p	p	root	nucleus	f

Appendix B.

Table 1. Annotation of the *Navy* excerpt. Information and pragmatics

NAVY (1)			**LEVEL B: information / pragmatics**				
			B1 information	**B2 illocution**		**B3. information patterning**	**B4 REF-unit**
SPEAKER	**TURN**	**SEGMENT**	**Functions**	**B2.1 illocution**	**B2.2 meta-illocution**	**Patterning**	**Type**
TOC	1	so I went to this &m	i-COB	0		subpattern	stanza
TOC	1	what I thought was my friend	PAR	0			
TOC	1	this navy captain down &dn	SCA	0			
TOC	1	at naval headquarters	COB	narration			
TOC	1	I said	INT	0		r-sub-patern	
TOC	1	this is terrybly akward	COM-r	reported-speech	protest		
TOC	1	I've just been promoted	COB-r	reported-speech	protest	r.sub-pattern	
TOC	1	from third mate	CMM-r	reported-speech	ill-pattern-comparison		
TOC	1	to second mate	CMM-r	reported-speech			
TOC	1	and [/1] and	AUX r	0			
TOC	1	could we	i-COM-r	0			
TOC	1	possibly	PAR	0			

Table 2. Annotation of the *Navy* excerpt. Information and pragmatics (2)

NAVY			**LEVEL B: information / pragmatics**				
			B1 information	**B2 illocution**		**B3. information patterning**	**B4 REF-unit**
SPEAKER	**TURN**	**SEGMENT**	**Functions**	**B2.1 illocution**	**B2.2 meta-illocution**	**Patterning**	**Type**
TOC	1	postpone these orders	COM-r	reported-speech			
TOC	1	for a little bit	APC	0	proposal		
TOC	1	my friend	TOP	0		sub-pattern	stanza
TOC	1	stood up	COB	narration			
TOC	1	behind his desk	SCA-f	0			
TOC	1	in his full	COB	narration			
TOC	1	&f four stripes	SCA-f	0			
TOC	1	and said	INT	0		r-sub-pattern	
TOC	1	Lieutenant	COM-r	reported speech	recall		
TOC	1	you	TOP-r	0		r-subpattern	
TOC	1	are in the US Navy now	COM-re	reported speech	alert		

Appendix C.

Table 1. Annotation of the *Navy* excerpt. Syntax and semantics (1)

NAVY			LEVEL C: cognition /syntax			
			C1 semantic interpretation	**C2 modality**	**C3 Phrase Structure**	**C4 syntactic compositionality**
SPEAKER	**TURN**	**SEGMENT**	**s-proiection**	**Modal change**	**Filling of the unit**	**Compositional result**
TOC	1	so I went to this &m		mA	incomplete S	
TOC	1	what I thought was my friend	p	mE	NP	
TOC	1	this navy captain down &dn	p	mA	NP	S
TOC	1	at naval headquarters	p	mA	PP	
TOC	1	I said	p	mA	S	
TOC	1	this is terrybly akward	p	mE	S	
TOC	1	I've just been promoted	p	mA	S	S
TOC	1	from third mate	n	mA	PP	
TOC	1	to second mate	p	mA	PP	
TOC	1	and [/1] and	n	0	con	
TOC	1	could we	n	0	AUX	
TOC	1	possibly	p	mE	ADV	
TOC	1	postpone these orders	p	mD	VP	S
TOC	1	for a little bit	p	0	PP	
TOC	1	my friend	p	mA	NP	
TOC	1	stood up	p	0	VP	VP

Table 2. Annotation of the *Navy* excerpt. Syntax and semantics (2)

NAVY			LEVEL C: cognition / syntax			
			C1 semantic interpretation	**C2 modality**	**C3 Phrase Structure**	**C4 syntactic compositionality**
SPEAKER	**TURN**	**SEGMENT**	**s-proiection**	**Modal change**	**Filling of the unit**	**Compositional result**
TOC	1	behind his desk	p	mA	PP	
TOC	1	in his full	n	0	incomplete PP	
TOC	1	&f four stripes	p	mA	NP	
TOC	1	and said	p	mA	S	
TOC	1	Lieutenant	p	mD	NP	
TOC	1	you	p	mA	NP	S
TOC	1	are in the US Navy now	p	mA	VP	

Appendix D.

Table 1. Annotation of the *Hearts* excerpt. Prosody

HEARTS			LEVEL A: prosody							
			A1 prosodic break	**A.2 break features**			**A3 prosodic unit features**			
SPEAKER	**TURN**	**SEGMENT**	**Break type**	**A2.1 reset**	**A2.2 lengthning**	**A2.3 pause**	**A2.4 int-drop**	**A3.1 unit-type**	**A3.2 unit structure**	**A3.3 Perceptively relevant F0-movement**
DAN	1	neat	//	p	p	p	p	root	nucl	r/platform
DAN	1	wait	//	n	n	p	p	root	nucl	platform
DAN	1	play novice	//	p	n	n	n	root	nucl	f/platform
DAN	1	I've never played Hearts	/	p	n	n	p	root	nucl	platform
DAN	1	before in my life	//	tc	tc	tc	tc	suffix	unstructured	flat
JEN	2	you've never played hearts	//	tc	tc	tc	tc	root	prep-nucleus	r
DAN	3	no	/	pmin	n	n	n	chained-root	nucl	platform
DAN	3	I don't know how to play it	//	tc	tc	tc	tc	chained-root	nucl	platform
JEN	4	oh	//	p	n	p	p	root	nucl	f
JEN	4	okay	/	p	n	n	n	chained-root	nucl	f
JEN	4	I'll teach you	//	tc	tc	tc	tc	chained-root	nucl	r/f
DAN	5	passing disabled	//	p	n	n	nb	root	nucl	f
DAN	5	that's you	//	tc	tc	tc	tc	root	nucl	r/f
JEN	6	queen of &sp	+					[0]		

Appendix E.

Table 1. Annotation of the *Hearts* excerpt. Information and

HEARTS pragmatics			**LEVEL B: cognition/information**			
			B1 information	**B2 illocution**	**B3**	**B4 REF-Unit**
SPEAKER	**TURN**	**SEGMENT**	**Functions**	**B2.1 illocution**	**Patterning**	**Type**
DAN	1	neat	COM	move (assent)	simple	UTT
DAN	1	wait	COM	move (waiting request)	simple	UTT
DAN	1	play novice	COM	assertion (self-conclusion)	simple	UTT
DAN	1	I've never played Hearts	COM	assertion (explanation)	complex	UTT
DAN	1	before in my life	APC	[0]		
JEN	2	you've never played hearts	COM	aknowledgment	simple	UTT
DAN	3	no	CMM	explication+reinforce	illocutionary pattern	CMM-UTT
DAN	3	I don't know how to play it	CMM	explication+reinforce		
JEN	4	oh	COM	expressive	simple	UTT
JEN	4	okay	CMM	conclusion+reinforce	illocutionary pattern	CMM-UTT
JEN	4	I'll teach you	CMM	conclusion+reinforce		
DAN	5	passing disabled	COM	citation (reading)	simple	UTT
DAN	5	that's you	COM	directive (agreement)	simple	UTT
JEN	6	queen of &sp	EMP	[0]	0	[0]

Appendix F.

Table 1. Annotation of the *Hearts* excerpt. Syntax and

HEARTS semantics			Level C Cognition / syntax			
			C1. semantic interpretation	C2 modality	C3 Phrase structure	C4 Syntactic compositionality
SPEAKER	TURN	SEGMENT	s-proiection	Modal change	Filling of the unit	Compositional result
DAN	1	neat	p	mE	AdgP	
DAN	1	wait	p	mD	VP	
DAN	1	play novice	p	mA	NP	
DAN	1	I've never played Hearts	p	mA	S	S
DAN	1	before in my life	p	0	PP	
JEN	2	you've never played hearts	p	mE	S	
DAN	3	no	p	mA	ADV	
DAN	3	I don't know how to play it	p	mE	S	
JEN	4	oh	n	m7	NT	
JEN	4	okay	p	m8	ADV	
JEN	4	I'll teach you	p	m8	S	
DAN	5	passing disabled	p	m9	NP	
DAN	5	that's you	p	m10	S	
JEN	6	queen of &sp	n	0	0	

CHAPTER 8

Comparing annotations for the prosodic segmentation of spontaneous speech

Focus on reference units

Alessandro Panunzi[i], Lorenzo Gregori[i] and Bruno Rocha[ii]
[i]University of Florence – LABLITA / [ii]Federal University of Minas Gerais

This chapter reports a quantitative and qualitative comparison of seven annotations performed on the same two American English texts: a monologue and a dialogue. The analysis of these data is complex, since the annotations have been made independently by each research group on the basis of their own theoretical frameworks. Despite this difference, the fundamental role of prosody in the analysis of speech emerges clearly in every annotation. Prosodic breaks can be then viewed as theory independent entities. After summarizing the key features of theoretical models, we derived a unified tagset and developed a web application (SLAC) to compare different annotations. Finally, agreement on prosodic breaks has been measured in different ways, reporting promising results in terminal break identification.

Keywords: spoken language models, segmentation criteria, prosodic breaks, annotation comparison, agreement

1. Introduction

In this chapter we will illustrate the results of the comparison between the seven segmentations and the corresponding analyses of the same American English texts extracted from the Santa Barbara Corpus (Du Bois et al., 2000–2005): *Navy* (a monologue) and *Hearts* (a dialogue).

As we have already said in the introduction to Part II of this book, we will refer to the annotations stored in the SLAC online database with the following abbreviations: CHA for Chafe; CNR for Cresti-Raso; IZR for Izre'el; KKP for Kibrik-Korotaev-Podlesskaya; MRT for Martin; MAY for Maruyama; and MIT for Mithun. Within the comparison of the theoretical models, we will refer to the papers in the previous chapters of Part II of the book as Cresti & Moneglia; Izre'el;

https://doi.org/10.1075/scl.94.17pan

Kibrik et al.; Martin; Maruyama; Mithun; Raso et al. (without further specifications). In Section 2 differences among the presented theoretical models are discussed through a comparative analysis on the following topics: (1) the segmentation of the speech flow into discrete units; (2) the nature of the reference unit of spoken language; (3) the relation between prosodic and syntactic structures. Section 3 presents the SLAC database and focuses on the description of the web query interface and on the Unified Tagset that has been derived to perform a comparative analysis. Data about annotation agreement are described and analyzed in Sections 4 and 5: two types of agreement have been calculated, overall and pairwise agreement on prosodic break identification.

2. Comparing the different theoretical perspectives

2.1 Preliminary remarks

In order to set the basis for the following discussion, it is useful to make some preliminary remarks on topics that will be dealt along this section. One of them is the concept of the reference unit of spoken language. Without claiming to be exhaustive about the subject, the reference unit could be defined as *a minimal unit of complete and autonomous communicative meaning that composes a spoken text* (Cresti, 2000; see also Barbosa & Raso, 2018; Moneglia & Raso, 2014). This comprehensive definition, however, should not obfuscate the different views regarding the nature of the communicative meaning conveyed by the reference unit. As explained in Section 2.4, some of the authors in this volume state that spoken language is organized in units that express a focus of consciousness of the speaker (Chafe, 1994), while others sustain that the reference units of spoken language corresponds to actions conveyed by the speaker toward the listener (Austin, 1962; Cresti, 2000; Moneglia & Raso, 2014).

It is largely recognized that prosody, along with syntax, is a core element that helps us to understand the structure of the spoken language and, for this reason, to set limits between its reference units. A more syntactically oriented approach considers prosody as subordinated to syntax, and the limits between reference units correspond to *theorecically proposed boundaries*. Prosodically oriented frameworks, on the other hand, will conceive the speech flow as a sequence of tone units separated one from another by prosodic boundaries, in such a way that the reference units of spoken language are segmented by *physically perceivable boundaries*.

It is important to notice that the syntactic criterion does not exclude the prosodic one and vice versa. In fact, in different frameworks, the reference units, "regardless of how they are defined, are separated by boundaries that are defined by highlighting greater or lesser perceptual or theoretical grounds" (Barbosa & Raso, 2018, p. 1364). The framework adopted by Maruyama, for instance, which is syntactically oriented, considers that the limits between reference units coincide with syntactic boundaries, but also it takes prosodic information into account to identify the internal divisions of reference units.

The framework adopted by Izre'el, which is prosodically oriented, rely mostly on the presence of acoustic cues to identify the limits between reference units. Even so, the author draws attention to two kinds of exceptions that denotes some theoretical ground on the identification of reference units: cases in which a reference unit ends with a minor prosodic boundary (such as some types of greetings) and cases in which it continues after a major prosodic boundary (such as discourse markers followed by conclusive boundaries). Even Raso et al., who make an effort to rely exclusively on prosodic information to segment the speech into terminated sequences, recognizes that the reference units are identified not only by prosodic cues, but also by its illocutionary properties.

2.2 The segmentation of the speech flow into discrete units

Given the aforementioned premises, we can assert that all annotators recognize that prosody plays an important role (by itself or along with syntax) in segmenting the speech flow. However, there are different approaches to explain how prosody works on the segmentation of discrete units.

In their respective chapters, Izre'el, Cresti & Moneglia, and Raso et al. declare that the speech flow is segmented by prosodic boundaries (or prosodic breaks), which can be divided on at least two types: terminal (conclusive, major boundaries) and non-terminal (continuative, minor boundaries). While terminal breaks signal that a sequence has prosodically reached to an end, non-terminal breaks signal the limits between two or more units that are part of the same terminated sequences. All teams recognize that prosodic breaks are complex phenomena and are due to a sum of factors such as "pause, f0 parameters, duration parameters, intensity parameters, rhythmic parameters, and voice quality change" (Raso et al.).

The presence of a prosodic break that segments the speech flow is easily perceived by both expert and non-expert annotators, as empirically shown by Raso et al.'s work. On the other hand, the ability to assign the value (terminal or non-terminal) to a break may depend on the expertise of the annotator. Terminal breaks have much higher inter-rate agreement than non-terminal ones, which could indicate the existence of more than one type of non-terminal break, varying in degree.

In the view described so far, prosodic breaks seem to be complex sets of prosodic cues that are intentionally added to certain parts of the speech flow in order to create chunks of information, indicating whether different chunks compose a single terminated sequence or belong to different ones. Prosodic breaks are always part of the prosodic contours and superpose on the segmental and suprasegmental level to other properties (illocutionary prosodic properties and attitudinal prosodic properties, for instance). Nevertheless, it seems that they exist as an independent entity on the conceptual level.

Some authors have declared that they segment the speech based only on prosodic cues that are present on the acoustic signal, and only later proceed to analyze the units that emerge from this segmentation. This is the case of the Izre'el, Cresti & Moneglia, Raso et al., Kibrik et al., and Mithun.

However, the segmentations by Kibrik et al. and Mithun are based on a criterion that might be slightly different from the previous one: the alternation of prosodic contours. As Mithun (this volume, Part II) says, in English, basic intonation units feature an initial pitch reset followed by declination. Because of that, the end of an intonation unit contrasts with the beginning of the following unit, making it possible to understand where the limits between them are located. Also, intonation units can be followed by pauses, which can be seen as an additional criterion to identify their limits. Kibrik et al. seem to have a similar view, since EDUs (Elementary Discourse Units) can be identified on the basis of the following prosodic properties: pausing (inhalation), primary accent (accentual center), integral tonal contour, tempo pattern, loudness pattern (Kibrik, Korotaev, & Podlesskaya, this volume, Part II).

Maryuama adopts a different procedure to segment the speech flow, based on the recognition of syntactic-discursive boundaries and also of prosodic boundaries and pauses (seen as different entities). LUUs (*Long Utterance Units*, the reference units of spoken language, understood as syntactic, discourse and interactional units in which the interaction between two speakers is based) are identified by syntactic boundaries (which tend to coincide to prosodic breaks), while the SUU (*Short Utterance Units*, which are internal divisions of LUUs) are identifiable only by prosodic cues.

Martin follows "the extended *Approche Pronominale*" (Blanche-Benveniste, 1990; Debaisieux, 2013; Deulofeu, 2003), in which the speech flow is segmented in prosodic and syntactic entities. The prosodic segmentation does not always correspond to the syntactic one, and because of that they should be carried out independently. On the prosodic level, the minimal speech units are stress groups, defined as sequences of syllables with only one stressed syllable (Martin, this volume, Part II). Each stress group ends with a distinctive prosodic contour that establishes a relation of dependency with the following group, creating a hierarchical relation between prosodic units in the speech. Every language has its own prosodic patterns, which can be more or less similar to the patterns of other languages. The syntactic level should be analyzed in two different subcomponents: microsyntax and macrosyntax. The first one accounts for the traditional syntactic relations between groups of words, phrases and clauses inside a given utterance. The macrosyntactic subcomponent correspond to discursive and/or pragmatic relations between units, just like the ones present in topic constructions such as *la coque, le sable est à quarante mètres* 'the shell, the sand is forty meters deep' (adapted from Debaisieux & Martin, this volume, Part I). Also, it should be noticed that macrosyntactic subcomponent regards not only a relation between constituents (microsyntactic units), but also between constituents and interjections, gestures, etc.

2.3 The relation between prosody and syntax

In the linguistic literature there is a great discussion on how the prosodic level interacts with the syntactic one. It is possible to discuss this relation from very different and complex perspectives, but we will focus on the existence of syntactic relations between:

1. The words and phrases that are located inside a tone unit;
2. Different tone units that compose a terminated sequence (i.e., a set of tone units concluded by a terminal prosodic break);
3. Different utterances.

In what concerns the first point, there seems to be consensus among all annotators in this volume that there is always compositionality on the syntactic level between words and phrases inside a tone unit. On the other hand, there is no complete agreement regarding the existence of syntactical relations between the tone units that compose a terminated sequence. Maruyama (this volume, Part II), for instance, defines LUU (*Long Utterance Units*) as units of syntactic, discursive and interactional nature, whose limits are signaled by syntactic properties and, in most of the cases, coincide with prosodic disjunctures. A LUU can be formed by a single or more than one SUU (*Short Utterance Unit*), separated by prosodic boundaries.

Mithun, Chafe, Izre'el, and Kibrik et al. seem to converge on the idea that the prosodic units of the same terminated sequence form a larger syntactic structure. Their basic assumption seems to be that terminal prosodic boundaries set the limits of units in which syntactic relations may occur, even though sometimes a mismatch between the prosodic and the syntactic structures can be noticed.

For Cresti and Moneglia, there is syntactic compositionality between some of the units that form an illocutionary pattern, as emerges from their annotation schema (particularly, comparing the column *B3 – information patterning* –to the column *C4 – syntactic compositionality*; see Appendices B, C, E and F in Cresti & Moneglia), but there are also some information units, like Parentheticals, that are not compositional in respect of the rest of the terminated sequence.

Martin considers that compositionality between units separated by prosodic boundaries is possible, but not mandatory. In fact, the author insists on the independency between syntactical and prosodic structures, which are seen as different resources available to the speaker in the text construction process (Debaisieux & Martin, this volume, Part I).

Not every author made comments on the existence of syntactic relations between utterances, but even so it is possible to make some observations. Izre'el retains that usually there are no relations between two different terminated sequences, but, in some cases, a phrase in a given terminated sequence can combine with elements of another one in order to compose a single structure. In the Cresti and Moneglia annotation schema, cases like that are regarded as independent syntactic structures. For Martin, conclusive contours indicate a macrosyntactic nucleus (a syntactic unit that is semantically autonomous, can form by itself an utterance and carries an illocutionary force). Even so, there can be some configurations in which two or more nuclei (that are not separated by conclusive contours) form more complex macrosyntactic patterns.

2.4 The nature of the reference units for spoken speech

The different annotation schemes reflect two basic tendencies to define the reference units of spoken speech. One of them, markedly influenced by Chafe's framework, would be to consider that spoken language is organized in sequences of intonation units, each one expressing a new idea and one focus of consciousness (Chafe, 1994). These cognitive units could be divided on at least two different types: substantive (units that present referents and events) and regulative (units that regulates the interaction). Also, an intonation unit can occur by itself, with a conclusive profile, or in combination with other intonation units, forming an intonation phrase.

Chafe, Mithun, Izre'el, Maruyama, and Kibrik et al. adhere to this view, adopting categories inspired by Chafe's framework to classify and describe intonation units. Among these authors, there is a tendency to consider *intonation units* (or *units of idea*, understood as their functional counterpart) as the reference unit for spoken language, which is precisely the case of Chafe and Mithun. For Maruyama, the reference units are defined as the segments delimited by LUUs (Long Utterance

Units) and SUUs (Short Utterance Units) boundaries, which constitute units of idea. For Kibrik et al., the reference units are the EDUs (Elementary Discourse Units), that represents a focus of conscience (Kibrik et al., this volume, Part I). For Izre'el, however, the reference unit is the *utterance*, defined as a set of *information modules* (which are units of idea).

The other tendency, present in works influenced by Language into Act Theory (Cresti, 2000; Moneglia & Raso, 2014), is to conceive that spoken language is organized in prosodically terminated sequences by which speakers convey speech acts. These sequences, which can be made of a single illocutionary unit (*utterance*) or a combination of more than one illocutionary unit (*stanza*), are seen as the reference units of speech due to their illocutionary properties. Both utterances and stanzas can have optional information units that do not convey illocutions, but regulate the interaction with the listener (*dialogic units*) or provide additional textual and pragmatic information in order to properly interpret the illocution (*textual units*). Two main aspects of this approach are: (1) the recognition of a necessary pragmatic level of communication and (2) the hierarchy between illocutionary and non-illocutionary units inside utterances and stanzas. Both Cresti and Raso et al. explicitly adopt this framework. However, it is worth noticing that Kibrik et al. also understand that each EDU conveys a speech act.

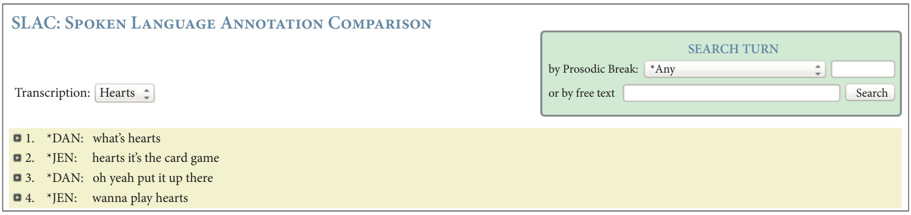
Martin works under similar assumptions. Debaisieux and Martin (this volume, Part I) define *utterances* as combinations of words that form a syntactical frame and convey an illocution by an appropriate prosodic contour. Also, the authors emphasize that a discourse can be fully appreciated only if the researcher does not conceive the communication just as sequences of textual units, but as sequences that can combine both textual and non-textual messages in macrosyntactic structures.

3. The SLAC database

3.1 Web interface

SLAC is a web tool designed to compare different prosodic annotations performed on the same texts, and specifically *Hearts* and *Navy*.[1]

[1] Freely accessible online at <http://lablita.it/app/slac>



SLAC: Spoken Language Annotation Comparison

SEARCH TURN

by Prosodic Break: *Any

or by free text

Transcription: Hearts

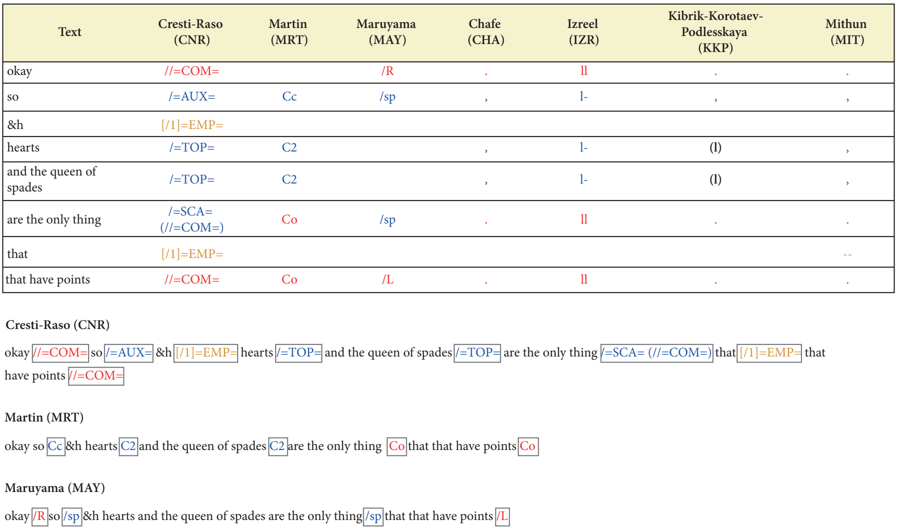
1. *DAN: what's hearts
2. *JEN: hearts it's the card game
3. *DAN: oh yeah put it up there
4. *JEN: wanna play hearts

Figure 1. The SLAC database web interface

In the access page (Figure 1) the full text transcription is displayed in a CHAT-like format (MacWhinney, 2000) without any annotations, divided in dialogic turns (the only type of segmentation that is free from any interpretation). Each line contains the turn number, the speaker identifier, the text of the turn transcription and the original audio fragment of the turn. A search function is also available, allowing users to search by free text or by prosodic break. In this case the matching turns of every annotations are displayed.

Prosodic annotation comparison can be performed by selecting one turn; here the text with multiple annotations is displayed in two ways: a tabular view, that highlights differences and similarities among annotations, and an inline view, in which the original annotations are displayed independently (Figure 2).

In the tabular view, transcription text is placed in the first column and an additional column is shown for each annotator. The table is filled with the tags put by each annotator in each text segment; if a text point is not marked by an annotator



Text	Cresti-Raso (CNR)	Martin (MRT)	Maruyama (MAY)	Chafe (CHA)	Izreel (IZR)	Kibrik-Korotaev-Podlesskaya (KKP)	Mithun (MIT)
okay	//=COM=		/R	.	ll	.	.
so	/=AUX=	Cc	/sp	,	l-	,	,
&h	[/1]=EMP=						
hearts	/=TOP=	C2		,	l-	(l)	,
and the queen of spades	/=TOP=	C2		,	l-	(l)	,
are the only thing	/=SCA= (//=COM=)	Co	/sp	.	ll	.	.
that	[/1]=EMP=						--
that have points	//=COM=	Co	/L	.	ll	.	.

Cresti-Raso (CNR)

okay //=COM= so /=AUX= &h [/1]=EMP= hearts /=TOP= and the queen of spades /=TOP= are the only thing /=SCA= (//=COM=) that [/1]=EMP= that have points //=COM=

Martin (MRT)

okay so Cc &h hearts C2 and the queen of spades C2 are the only thing Co that that have points Co

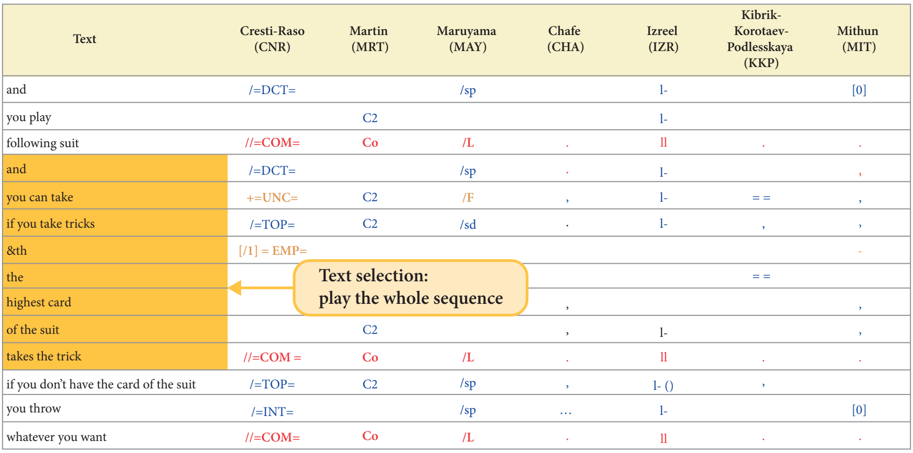
Maruyama (MAY)

okay /R so /sp &h hearts and the queen of spades are the only thing /sp that that have points /L

Figure 2. SLAC interface: tabular and inline view

the table cell is blank. For example, Figure 2 shows that a prosodic break is perceived by most of the annotators after *okay* (line 1), but not by Martin, that didn't put any tag after that token.

The number of rows depends on the number of annotation tags: there is a line break if at least one annotator placed a tag. For example, nobody perceived a break within the token sequence *and the queen of spades*, that is displayed on the same line (line 5); on the contrary there is a line break after the token *&h* (line 3) because Cresti & Raso (and no one else) put a tag. Audio speech is directly available from the table. It is possible to play the audio of a single text chunk or of a sequence of more chunks (by mouse dragging, as in Figure 3).



Text	Cresti-Raso (CNR)	Martin (MRT)	Maruyama (MAY)	Chafe (CHA)	Izreel (IZR)	Kibrik-Korotaev-Podlesskaya (KKP)	Mithun (MIT)
and	/=DCT=		/sp		l-		[0]
you play		C2			l-		
following suit	//=COM=	Co	/L	.	ll	.	.
and	/=DCT=		/sp	.	l-		,
you can take	+=UNC=	C2	/F	,	l-	= =	,
if you take tricks	/=TOP=	C2	/sd	.	l-	,	,
&th	[/1] = EMP=						-
the						= =	
highest card				,			,
of the suit		C2		,	l-		,
takes the trick	//=COM =	Co	/L	.	ll	.	.
if you don't have the card of the suit	/=TOP=	C2	/sp	,	l- ()	,	
you throw	/=INT=		/sp	...	l-		[0]
whatever you want	//=COM=	Co	/L	.	ll	.	.

Figure 3. SLAC interface: text-to-speech alignment

Behind this basic interface, SLAC data are stored in a database, that can be queried to derive global and local statistics about agreement and disagreement between different annotations. In order to allow the comparison, we developed a Unified Tagset (see Section 3.2 and the Appendix for more details) that classifies tags in five main classes, reported in the SLAC interface and displayed with different tag colors and formats:

1. Terminal breaks in red and bold;
2. Non-terminal breaks in blue;
3. Interruptions and disfluences in orange and bold;
4. Units which have not been transcribed by one annotator are marked with [ntrsc] and are displayed in black;
5. Alternative tagging is displayed in brackets.

3.2 The Unified Tagset

Each annotation uses a specific tagset designed for representing different aspects of speech. All of them deal with the prosodic level, while some of them includes also syntactical and discourse level tags. We have chosen to focus the Unified Tagset on the prosodic level and, more specifically, on the difference between terminal and non-terminal prosodic boundaries. Syntactic tags were not included, just as prosodic tags that did not signal the presence and type of the prosodic boundaries. We have also included symbols that marked the presence of a disfluency that creates a prosodic boundary between units.

The main goal of the tagset is to provide an annotation schema that is *unified* and *specific* at the same time. By unified, we mean that the schema should represent the difference between terminal and non-terminal boundaries, which are notions present in most of the works, as well as the annotation of prosodic disfluencies on the limits of prosodic units. On the other hand, while doing this, we want to preserve the original tags used by the authors, in order to interfere the minimum as possible on their annotation. The solution proposed here is to report the different tags used by each annotation to the distinction between terminal (marked in red and bold) and non-terminal boundaries (marked in blue), even if it is not explicitly stated by each author.

For instance, we considered Maruyama distinction between LUUs and SUU as corresponding to the distinction between (respectively) terminal and non-terminal boundaries, since (1) Den et al. (2010) claims that LUUs tend to coincide with conclusive prosodic boundaries, other than with a syntactical and discourse level boundary; (2) comparing the annotations provided by the Japanese team with the other annotations, it can be observed that LUUs mostly coincide with terminal boundaries marked by other teams, and SUUs tend to coincide with non-terminal boundaries.

For what regards the annotation by Kibrik et al., it is possible to note the distinction between conclusive and continuative boundaries of the EDUs on their annotation schema: a group of punctuation marks signals transitional-continuity properties, while other signals illocutionary force. Illocutionary punctuation marks can be used in combination with transitional-continuity ones, but when they are used alone they convey a conclusive value to the sequence.

In Mithun and Chafe annotation, it is possible to notice some cases in which intonation contours are not delimited by punctuation marks. In this case, we considered that the absence of punctuation indicates that the boundary is marked by the presence of a pause and not by an explicit continuation tone of the prosodic contour. Because of that, we added the symbol [0], that indicates a non-terminal boundary. The Unified Tagset is reported in both the Appendix of this paper and on the SLAC website.

4. Overall agreement

4.1 Starting data and preliminary choices

Starting from data collected in the SLAC database, we carried out a comparative analysis of the seven different annotations of the full texts. Each annotation team started from the raw texts reported at the end of the introduction to the second part of the book.

First of all, it has to be underlined that the comparison does not aim to measure a parametrized inter-annotator agreement (e.g., the k-coefficient by Carletta, 1996). As a matter of fact, the annotation teams performed a segmentation task on the same texts, but each one used a different theoretical framework and a different set of tags, which have been merged *a posteriori* in the Unified Tagset. Although we will present a tentative application of standard agreement measures to the task of break identification in Section 4.3, the main focus of our work is to compare the results of different procedures applied on the same texts.

The specific aim of this comparison is then to measure how much different annotations performed with different perspectives can agree with respect to the segmentation of a spoken text, given that all these perspectives share the assumption of the fundamental role of prosody in structuring the spoken language. Before entering into the analysis, we have to introduce some essential preliminary remarks.

First of all, the contrastive analysis relies on the union of all possible segmentation units marked in each annotation. In other words, we have considered the positions in which at least one annotator marked a terminal or non-terminal break. Second, we excluded from the analysis: (1) all the units in which at least one annotator marked an interruption or a disfluency; (2) all the units which have been not transcribed (tag [ntrsc]) by at least one annotator. In order to evaluate the consensus of annotations from a stricter point of view, we took into account only the units which do not occur at the end of turn. This is motivated by the obvious fact that the turn change "forces" the annotation of a (tendentially terminal) break: considering the segmentation at the end of turn will automatically rise the consensus, introducing a bias in the measurement of how much the annotations really agree. Tables 1 and 2 report the number of terminal and non-terminal breaks for each annotation, respectively in *Hearts* and *Navy*, excluding the above-mentioned cases.

Table 1. Number of terminal and non-terminal breaks in *Hearts*

Hearts	CNR	MRT	MAY	CHA	IZR	KKP	MIT
Terminal	24	33	38	34	34	26	29
Non-Terminal	36	23	13	27	30	28	36
Total	**60**	**56**	**51**	**61**	**64**	**54**	**65**

Table 2. Number of terminal and non-terminal breaks in *Navy*

Navy	CNR	MRT	MAY	CHA	IZR	KKP	MIT
Terminal	9	16	17	10	13	6	9
Non-Terminal	45	20	24	21	33	23	33
Total	**54**	**36**	**41**	**31**	**46**	**29**	**42**

We organized our starting data distinguishing between two main cases. The first case comprehends all units for which at least one annotator marked a terminal break. Data are showed in Table 3; in the TB,NTB columns, values at the left of the comma represent the number of annotations in which there is a terminal break (TB), while values at the right of the comma correspond to the annotation that marked a non-terminal break (NTB). For instance, score 7,0 means that seven annotations on seven record a terminal break; score 5,2 means that five annotations record a terminal break and two a non-terminal break; score 2,1 means that two annotations record a terminal break, one records a non-terminal break (and four do not record any mark). The second and third colums respectively report the number and the percentage of breaks in which we found the agreement score indicated by the TB,NTB column.

Table 3. Overall break annotation in the positions where at least one annotator put a terminal break (in *Navy* and in *Hearts*)

Hearts			***Navy***		
TB,NTB	**n**	**%**	**TB,NTB**	**n**	**%**
7,0	15	31.91%	7,0	3	12.50%
6,1	6	12.77%	6,1	2	8.33%
6,0	1	2.13%	5,2	2	8.33%
5,2	7	14.89%	4,3	2	8.33%
4,2	2	4.26%	4,2	2	8.33%
3,4	1	2.13%	3,3	1	4.17%
3,3	1	2.13%	3,1	1	4.17%
3,2	1	2.13%	2,5	2	8.33%
2,5	4	8.51%	2,3	1	4.17%
2,3	1	2.13%	2,1	1	4.17%
2,0	1	2.13%	1,5	1	4.17%
1,6	1	2.13%	1,4	1	4.17%
1,5	1	2.13%	1,3	1	4.17%
1,4	2	4.26%	1,2	3	12.50%
1,1	2	4.26%	1,1	1	4.17%
1,0	1	2.13%			
tot	**47**	**100.0%**	**tot**	**23**	**100.0%**

The second case we treated corresponds to all units for which at least one annotation records a non-terminal break, but no one marked a terminal one. Data are reported in Table 4; obviously, the values at the left of the comma in the columns TB,NTB in this cases is always 0.

Table 4. Overall break annotation in the positions where at least one annotator put a non-terminal break and no one put a terminal

Hearts			*Navy*		
TB,NTB	**n**	**%**	**TB,NTB**	**n**	**%**
0,7	1	2.2%	0,7	3	5.26%
0,6	3	6.67%	0,6	4	7.02%
0,5	4	8.89%	0,5	4	7.02%
0,4	5	11.11%	0,4	4	7.02%
0,3	5	11.11%	0,3	9	15.79%
0,2	9	20.00%	0,2	5	8.77%
0,1	18	40.00%	0,1	28	49.12%
tot	**45**	**100.0%**	**tot**	**57**	**100.0%**

Preliminary data show that the consensus is generally higher, in Tables 3 and 4, on the dialogical text (*Hearts*) than on the monological one (*Navy*); the same tendency is confirmed by all the results that emerged from the comparison (see Tables 5–14). These data were actually expected, since dialogical communicative events are based on the accomplishment of an interactive sequence of communicative acts, which are largely encoded by the prosody. Tonal segmentation is then highly prominent, in order to clearly specify to the listener the parsing of the speech flow into single acts, which are short and highly variated. On the contrary, the speaker of a monologue aims to express his own thought by means of a more complex text. In this kind of activity, the focus is on the text production rather than on the execution of single communicative acts. Prosodic segmentation tends to identify larger units, leaving space for a more syntactic organization of the speech, and it is therefore less prominent and identifiable.

Moreover, it is worth noticing that the consensus is always higher for the positions in which at least one annotation records a TB. As a matter of fact, the highest values in Table 3 is on total agreement row (value 7,0 records 31.91% in *Hearts* and 12.50% in *Navy*); conversely, Table 4 shows that, where nobody marked a TB, the total agreement (row 0,7) records the lowest values in both texts, and the number of cases in which only one annotation marked a NTB is very high (40% for *Hearts*, 49.12% for *Navy*).

For these reasons, and to minimize data sparseness, the set of units in which at least one annotator marked a terminal break have been assumed as the baseline for our further analyses. This means that in the following part of our analysis we will consider only the positions where at least one annotator put a terminal break, that is, all the cases described in Table 3.

4.2 Interpreting the data

Since data presented in Table 3 are too sparse to extract a general evaluation of the inter-annotation agreement, we grouped the results starting from two perspectives, each one focusing on a different aspect of the consensus rate. In the first evaluation (*ANY*), we considered both terminal and non-terminal signs as having the same value: both T and NT count as 1. For instance, lines corresponding to scores 7,0 / 6,1 / 5,2 / 4,3 have been valued at 7/7; lines corresponding to scores 5,0 / 4,1 / 3,2 / 1,4 have been valued at 5/7. Table 5 reports these data. In this evaluation, we focused on the following question: how much is the consensus about the presence of a break in all positions in which at least one annotation reports a terminal break?

Table 5. ANY: break consensus where at least one annotator put a terminal break

Agreement	***Hearts***		***Navy***		**Total**	
7/7	34	72.34%	11	45.83%	45	63.38%
6/7	5	10.64%	4	16.67%	9	12.68%
5/7	4	8.51%	2	8.33%	6	8.45%
4/7	0	0.00%	2	8.33%	2	2.82%
3/7	0	0.00%	4	16.67%	4	5.63%
2/7	3	6.38%	1	4.17%	4	5.63%
1/7	1	2.13%	0	0.00%	1	1.41%
	47		24		71	

From this table, it clearly emerges that the positions in which at least one annotation marked a TB have been well identified also by all others, at least with a NTB. Total agreement (7/7) is much higher in *Hearts* (72.34%) than in *Navy* (45.83%), but in both the text it records a higher percentage with respect to all the other cases of partial agreement (6/7, 5/7, etc.) reported in Table 5. We can conclude that the identification of units characterized by a TB is someway independent from the theoretical framework: even if the annotations started from different assumptions, the basic segmentation of a text is pretty constant in all cases.

In the second evaluation (*OTB*), we considered only terminal breaks: in this perspective, terminal breaks count as one, while non-terminal breaks are valued at 0. In other words, we simply did not consider the value at the right of the comma in the reported scores: 5,2 / 5,1 / 5,0 have been valued at 5, while 4,3 / 4,2 / 4,1 / 4,0 have been valued at 4. This evaluation is reported in Table 6, and tries to answer to the following issue: how much is the identification of terminal break shared between annotations that follow different criteria?

Table 6. OTB: terminal break consensus where at least one annotator put a terminal break

Agreement	*Hearts*		*Navy*		Total	
7/7	15	31.91%	3	12.50%	18	25.35%
6/7	7	14.89%	2	8.33%	9	12.68%
5/7	7	14.89%	2	8.33%	9	12.68%
4/7	2	4.26%	4	16.67%	6	8.45%
3/7	3	6.38%	2	8.33%	5	7.04%
2/7	6	12.77%	4	16.67%	10	14.08%
1/7	7	14.89%	7	29.17%	14	19.72%
	47		24		71	

Before moving to some considerations about this data, we would like to stress that a high *OTB* agreement on seven different annotations is in principle very hard to obtain, especially if we exclude the units at the END of turns.

Data reported in Table 6 do not sketch a clear picture as the one in Table 5, and the differences among the two texts seem much more evident. In order to facilitate the analysis, we added another level of abstraction, separating the cases in four macroclasses. From these data, it emerges that strong disagreement is never very high. More specifically, it comes out that two thirds of the breaks in *Hearts* and half of the breaks in *Navy* record a total or at least strong agreement (see data in Table 7):

1. "Total agreement" value corresponds only to the cases in which seven annotations on seven marked a TB (value 7/7);
2. "Strong agreement" value merges the cases in which only one annotation on seven disagrees with respect to the others: it subsumes 6/7 (6 put the TB while 1 did not) and 1/7 (6 did not put the TB while only 1 did) values;
3. "Partial disagreement" value merges the cases in which just two annotations on seven disagree with respect to the others: it subsumes 5/7 (5 put the TB while 2 did not) and 2/7 (5 did not put the TB while 2 did) values;
4. "Strong disagreement" value merges the most problematic cases, in which almost half of the annotations disagree with the others: it subsumes 4/7 and 3/7 values.

Table 7. Summary of the overall agreement

	Hearts		*Navy*	
Total agreement (7/7)	15	31.91%	3	12.50%
Strong agreement (6/7 + 1/7)	14	29.79%	9	37.50%
Partial disagreement (5/7 + 2/7)	13	27.66%	6	25.00%
Strong disagreement (3/7 + 4/7)	5	10.64%	6	25.00%

4.3 Standard agreement

The use of standard measures is important to perform a more grounded data analysis, although the results must be interpreted within this reference frame and not as absolute agreement values. As we already noticed at the beginning of Section 4.1, the following analysis violates a basic assumption of agreement measurement, that is, the execution of the same task by all the annotators. On the contrary, here the annotation has been performed indepentently by seven teams, on the basis of different theoretical frameworks, and finally results have been normalized to match a segmentation with terminal and non-terminal breaks.

In order to measure the agreement on terminal break identification, we needed to determine the whole set of positions where the presence of a terminal break is possible. This is not a trivial task, and involves phonetic and phonological features of the words produced in the speech flow. The analysis has been made in two steps: first, the set of positions is identified on the written transcription only, then an assessment of values has been performed by three annotators with the audio track.

The analysis on written texts started with the identification of the words that are normally stressed (nouns, adverbs, adjectives, predicative verbs) and unstressed (prepositions, conjunctions, pronouns, copulative verbs). Positions have been specified with the following rules:

1. Two stressed words are different phonological words: every phonological word must have one and only one stressed syllable;
2. Unstressed words are part of a bigger phonological word and have been attached to the preceding stressed syllable: note that there is no agreement on the fact that unstressed words need to be considered as part of the following or preceding syllable.

In the second step, the speech flow has been analyzed by three annotators and the positions have been adjusted until a shared judgment was reached. This assessment with the audio track was necessary to obtain a validated result, also considering that in spoken English it is frequent that stressed words are pronounced as unstressed.

Agreement is measured on the two texts, *Navy* and *Hearts*, jointly (Table 8) and separately (Tables 9 and 10) by using the following standard measures:

1. Observed agreement (measured agreement without adjusting for chance);
2. Multi-k (Davies & Fleiss, 1982);
3. Multi-p (Fleiss, 1971)
4. Krippendorff's alpha (Krippendorff, 1980)

We measured the agreement on terminal and non-terminal break (Real), on ANY break and on terminal break only (OTB).[2] The agreement on Real breaks is measured on a three-classes classification (terminal break, non-terminal break, no break) and is based on the original assignments of breaks by each annotator in each position. Agreements on ANY and OTB are measured on a two-classes classification: both terminal and non-terminal breaks versus no break (ANY), terminal breaks versus both no break and non-terminal (OTB).

Table 8. Agreement measured on *Navy* and *Hearts*

Measure	Real	ANY	OTB
Observed Agreement	0.782	0.838	0.928
Multi-k	0.570	0.640	0.713
Multi-p	0.570	0.640	0.712
Krippendorff's alpha	0.570	0.640	0.713

Table 9. Agreement measured on *Navy*

Measure	Real	ANY	OTB
Observed Agreement	0.739	0.789	0.930
Multi-k	0.458	0.520	0.575
Multi-p	0.457	0.519	0.574
Krippendorff's alpha	0.457	0.519	0.574

Table 10. Agreement measured on *Hearts*

Measure	Real	ANY	OTB
Observed Agreement	0.816	0.876	0.926
multi-k	0.644	0.727	0.761
multi-p	0.644	0.727	0.761
Krippendorff's alpha	0.644	0.727	0.761

The agreement values with three different measures (multi-k, multi-p and Krippendorff's alpha) are nearly the same; this means that the dataset is homogeneous (data distribution is not affected by prevalence or bias). Data confirm the differences between monologue and dialogue observed before: a lower agreement in *Navy* and higher agreement in *Hearts*. Finally, the agreement OTB is always

[2] Standard agreement has been computed on the same dataset used for previous analysis, without the end of turns, the positions where at least one annotator marked an interruption or a disfluency and the ones without a full transcription (see Section 4.1 for details).

higher than ANY. This can appear inconsistent with the previous numbers (see Tables 5 and 6), but it actually depends on the fact that the two-classes are strongly unbalanced in OTB, where negative class includes both positions where annotators put a non-terminal break and the ones where they did not mark any break; this increases the agreement on the negative class. Otherwise ANY is less unbalanced, given that any break belong to the positive class and negative class contains only the positions where the annotator did not mark a break.[3]

These considerations highlight a low adequacy in adopting a standard measure to have a proper agreement estimation. For this reason, the conclusions drawn in previous and following paragraphs relies on a more grounded data analysis.

5. Pairwise agreement

The pairwise agreement is calculated between any pair of annotators, in order to perform a deeper analysis of the links between annotations that rely on different theoretical frameworks. The seven annotators produced 21 pairs to be compared.

As for the previous tasks, the starting point of this analysis is the terminal break table (Table 3), that is, the tables of terminal breaks from which interruptions and end of turns have been excluded. In fact these positions alter the agreement analysis in negative and in positive: interrupted units have a low relevance as reference segmentation units and end of turns have an obvious agreement on terminal breaks. According to the general analysis, we took into account only the positions where at least one annotator in the pair put a terminal break.

5.1 ANY: agreement on prosodic break perception

The pairwise agreement on *ANY* break measures the agreement about prosodic break perceptions: it does not consider the difference between terminal and non-terminal breaks.

The overall high agreement reached in this task validates prosodic break as a theory independent unit, given that its value is shared among different perspectives. We remark that, while the measures consider valid any break (T and NT), only the positions where at least one annotator put a T break are analyzed.

Given two annotators, *ANY* measure is defined as in Formula (1):

[3] Size of negative class on both texts: 85% of the positions belong to the negative class in OTB; 65% of positions belong to the negative class in ANY.

$$ANY = \frac{p(t,t) + p(t,nt)}{p(t,t) + p(t,nt) + p(t,\emptyset)} \tag{1}$$

where

1. *p(t,t)* is the number of positions where both annotators put a terminal break;
2. *p(t,nt)* is the number of positions where one annotator put a terminal break and the other one put a non-terminal break;
3. *p(t,∅)* is the number of positions where one annotator put a terminal break and the other one did not put any break.

Table 11. Pairwise agreement with ANY measure

Pair	**Agr. on *Hearts***	**Agr. on *Navy***	**Total agr.**
CNR-MRT	0.94	0.94	0.94
CNR-MAY	0.95	**1.00**	0.96
CNR-CHA	0.97	0.92	0.96
CNR-IZR	0.97	0.80	0.92
CNR-KKP	**1.00**	**1.00**	**1.00**
CNR-MIT	0.97	0.92	0.95
MRT-MAY	0.88	0.73	0.83
MRT-CHA	0.87	0.58	0.77
MRT-IZR	0.92	0.72	0.86
MRT-KKP	0.94	0.94	0.94
MRT-MIT	0.87	0.83	0.86
MAY-CHA	0.86	0.76	0.84
MAY-IZR	0.91	0.71	0.85
MAY-KKP	**0.98**	**1.00**	**0.98**
MAY-MIT	0.86	0.88	0.87
CHA-IZR	0.90	0.59	0.80
CHA-KKP	0.94	**1.00**	0.96
CHA-MIT	0.95	0.85	0.92
IZR-KKP	0.97	0.73	0.90
IZR-MIT	0.94	0.94	0.94
KKP-MIT	0.97	**1.00**	**0.98**
Average	**0.93**	**0.85**	**0.91**

Data are reported in Table 11. The overall agreement on both text is high: between 0.77 and 1.0. These results show that, while the upper bound of agreement is 1.0 for both texts, there is a greater variation in the lower bound: 0.86 for *Hearts* and 0.58 for *Navy*. This confirms that when the speech is more interactive, the agreement on break perception is higher; conversely monologues are less interactive and more subject to theory-driven interpretation.

5.2 OTB: agreement on terminal break

The agreement on terminal breaks (*OTB*) is based on the relation between the positions where both annotators agree on terminal breaks and the positions where they disagree. Given two annotators, *OTB* measure is defined as in Formula (2):

$$OTB = \frac{p(t,t)}{p(t,t) + p(t,nt) + p(t,\emptyset)} \tag{2}$$

where *p(t,t), p(t,nt)* and *p(t,∅)* are defined as in *ANY*.

Data are reported in Table 12.

Table 12. Pairwise agreement with OTB measure

Pair	**Agr. on *Hearts***	**Agr. on *Navy***	**Total agr.**
CNR-MRT	0.58	0.47	0.55
CNR-MAY	0.55	0.53	0.54
CNR-CHA	0.66	0.58	0.64
CNR-IZR	0.73	0.47	**0.65**
CNR-KKP	0.67	0.50	0.63
CNR-MIT	0.68	0.38	0.59
MRT-MAY	0.65	0.50	0.60
MRT-CHA	**0.76**	0.37	0.63
MRT-IZR	**0.76**	**0.61**	**0.71**
MRT-KKP	0.64	0.22	0.50
MRT-MIT	0.63	0.39	0.55
MAY-CHA	0.64	0.59	0.62
MAY-IZR	0.64	0.43	0.57
MAY-KKP	0.52	0.35	0.47
MAY-MIT	0.52	0.53	0.52
CHA-IZR	0.74	0.35	0.63
CHA-KKP	0.67	**0.60**	**0.65**
CHA-MIT	0.70	0.46	0.64
IZR-KKP	0.67	0.27	0.55
IZR-MIT	**0.75**	0.38	0.63
KKP-MIT	0.67	0.36	0.59
Average	**0.66**	**0.44**	**0.59**

The *OTB* between two annotators is medium-low: between 0.47 and 0.71. In these data strong differences emerge between *Navy* and *Hearts*. If we analyze the two texts separately we see that *OTB* is highly text-dependent. First of all the agreement range is very different: *OTB* in *Hearts* is between 0.52 and 0.76, while in *Navy* it is between 0.22 and 0.61. Then the pairs distribution is not uniform: for example Mithun-Izreel (IZR-MIT) have a relatively high agreement in *Hearts* (0.75) and a very low one in *Navy* (0.38).

5.3 Strong and weak disagreement

The *OTB* measure consider a disagreement if two annotators put two breaks, one terminal and one non-terminal, or if one annotator put a terminal break and the other put nothing. This is probably too rigid to make a fine-grained comparison and led us to distinguish between strong and weak disagreement:

1. Strong disagreement: *p(t,∅)*;
2. Weak disagreement: *p(t,nt)*;
3. Agreement: *p(t,t)*.

Table 13. Strong and weak disagreement in both *Hearts* and *Navy*

Pair	Agreement	Strong disagr.	Weak disagr.
CNR-MRT	0.55	0.13	0.88
CNR-MAY	0.54	0.08	0.92
CNR-CHA	0.64	0.12	0.88
CNR-IZR	**0.65**	0.24	0.76
CNR-KKP	0.63	0.00	**1.00**
CNR-MIT	0.59	0.11	0.89
MRT-MAY	0.60	0.42	0.58
MRT-CHA	0.63	**0.62**	0.38
MRT-IZR	**0.71**	**0.50**	0.50
MRT-KKP	0.50	0.11	0.89
MRT-MIT	0.55	0.32	0.68
MAY-CHA	0.62	0.43	0.57
MAY-IZR	0.57	0.36	0.64
MAY-KKP	0.47	0.03	**0.97**
MAY-MIT	0.52	0.28	0.72
CHA-IZR	0.63	**0.52**	0.48
CHA-KKP	**0.65**	0.13	0.88
CHA-MIT	0.64	0.22	0.78
IZR-KKP	0.55	0.22	0.78
IZR-MIT	0.63	0.16	0.84
KKP-MIT	0.59	0.06	**0.94**
Average	**0.59**	**0.24**	**0.76**

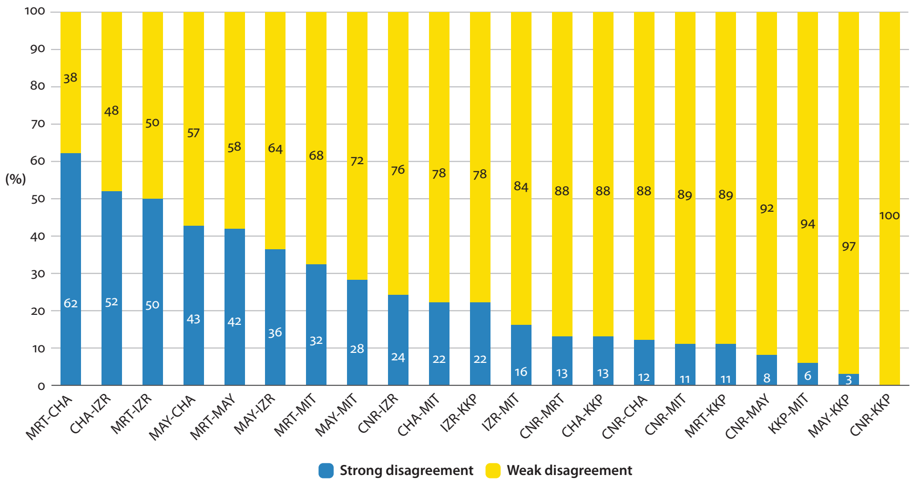


Figure 4. Strong and weak disagreement in both *Hearts* and *Navy*

Data (Table 13 and Figure 4) show that, in fact, for most of the pairs, the strong disagreement is less than 30% of the whole disagreement. Then there are five pairs in which the strong disagreement is relevant (30%–50%): MRT-MIT, MAY-IZR, MRT-MAY, MAY-CHA, MRT-IZR. Finally for two pairs the strong disagreement is preeminent (> 50%), that is, greater than the weak disagreement: MRT-CHA and CHA-IZR.

In addition to this, we notice that the distribution of strong disagreement does not correlate with the *OTB* values. For example Martin and Izreel (MRT-IZR) have the highest agreement on terminal break (0.75), but also their strong disagreement is very high (0.50); this means that the annotators often agree on terminal break, but in the other cases they disagree even on the presence of a break. Conversely Maruyama and Kibrik-Korotaev-Podlesskaya (MAY-KKP) have a very low OTB (0.47), but their strong disagreement is 0.03; in this case the two annotators often disagree about terminal breaks, but when they do it, they agree on the presence of a prosodic break.

5.4 Weighted agreement

Starting from these evidences a Weighted Agreement (*WA*) measure was defined, in order to take into account the difference between strong and weak disagreement. Given two annotators, *WA* measure is defined as in Formula (3):

$$WA = \frac{w_1 \cdot p(t,t) + w_2 \cdot p(t,nt)}{p(t,t) + p(t,nt) + p(t,\emptyset)} \qquad (3)$$

where

1. $w_1 = 1$;
2. $w_2 = 0.5$;
3. *p(t,t)*, *p(t,nt)* and *p(t,∅)* are defined as in *ANY*.

WA assigns a weight of 0.5 to the positions with weak disagreement and a weight of 0.0 to the ones with strong disagreement, so performing a better approximation to the real agreement between annotators.

Data about annotation pair similarities in *Navy* and *Hearts* according to *WA* highlight two properties of agreement (see Table 14). First of all the agreement is always higher in *Hearts* than in *Navy*, confirming the previous results on the correlation between agreement on terminal break identification and speech

Table 14. Pairwise agreement with WA measure

Pair	**Agr. on *Hearts***	**Agr. on *Navy***	**Total agr.**
CNR-MRT	0.76	0.71	0.75
CNR-MAY	0.75	**0.76**	0.75
CNR-CHA	0.81	**0.75**	**0.80**
CNR-IZR	**0.85**	0.63	0.78
CNR-KKP	**0.83**	**0.75**	**0.81**
CNR-MIT	0.82	0.65	0.77
MRT-MAY	0.77	0.61	0.72
MRT-CHA	0.82	0.47	0.70
MRT-IZR	**0.84**	0.67	0.79
MRT-KKP	0.79	0.58	0.72
MRT-MIT	0.75	0.61	0.71
MAY-CHA	0.75	0.68	0.73
MAY-IZR	0.77	0.57	0.71
MAY-KKP	0.75	0.68	0.73
MAY-MIT	0.69	0.71	0.70
CHA-IZR	0.82	0.47	0.71
CHA-KKP	0.81	**0.80**	**0.80**
CHA-MIT	0.82	0.65	0.78
IZR-KKP	0.82	0.50	0.73
IZR-MIT	**0.85**	0.66	0.79
KKP-MIT	0.82	0.68	0.78
Average	**0.79**	**0.65**	**0.75**

interactivity. The second property is the high variability in the pair agreement depending on the text: so we find pairs with high or low agreement in both *Navy* and *Hearts*, but also pairs with high agreement in the latter and low agreement in the former. These are some examples:

1. Martin and Mithun (MRT-MIT) have a low agreement in both texts;
2. Chafe and Kibrik-Korotaev-Podlesskaya (CHA-KKP) have a high agreement in both texts;
3. Izreel and Kibrik-Korotaev-Podlesskaya (IZR-KKP) have a high agreement in *Hearts* and a low agreement in *Navy*;
4. Maruyama and Mithun (MAY-MIT) have a high agreement in *Navy* and a low agreement in *Hearts*.

Figure 5 shows the pairwise agreement on a 2D map, where *x* and *y* axis are the agreement on *Navy* and *Hearts*, accordingly. The sparsity of points and the axis value spans are evidences of the aforementioned properties: agreement variability and higher agreement on *Hearts*. Besides this, the map highlights groups of annotators, according to their agreement in both texts.

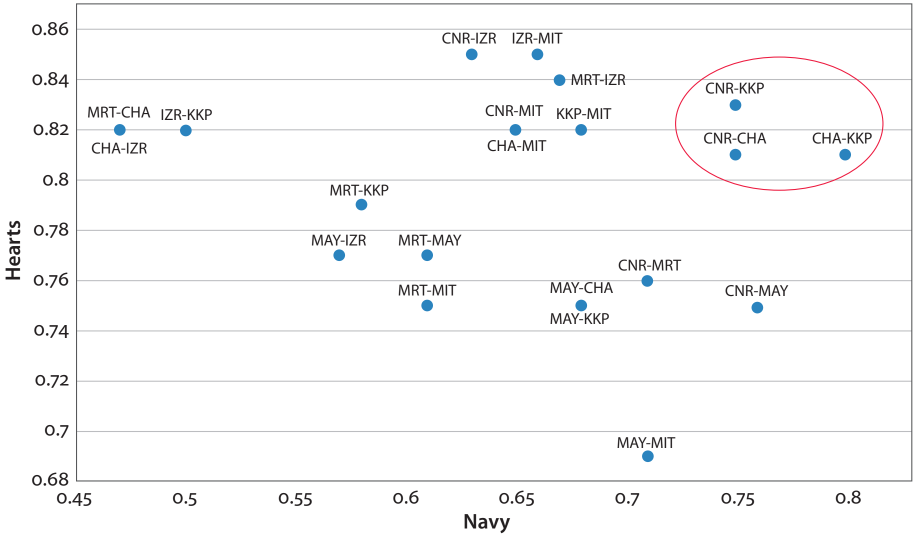


Figure 5. Pairwise agreement on a 2D map (x = agreement in *Navy*; y = aggrement in *Hearts*)

The group that mostly agree each other are Cresti-Raso (CNR), Chafe (CHA) and Kibrik-Korotaev-Podlesskaya (KKP). In fact, every pair among these annotators (3 elements) is placed in the top-right quarter, meaning that they have a high agreement on both texts (red circle in Figure 5). The pairwise comparison aimed to perform a deeper analysis on the agreement of different annotators from different perspectives. Three measures have been defined, *ANY, OTB* and *WA* and the main results can be summarized as follows:

1. Agreement on prosodic break identification is higher with more interactive dialogues;
2. Similarities between annotators on prosodic break identification is not dependent on the dialogic text type; conversely similarities on terminal break identification is strongly affected by the dialogue type;
3. There is a group of annotators with shared agreement on terminal break identification: Cresti-Raso, Kibrik-Korotaev-Podlesskaya and Chafe.

6. Final remarks

In the first part of this chapter, we conducted a qualitative comparison of the theoretical frameworks and the segmentation procedures adopted by each group of annotators in this volume (Section 2). We observed that there are different perspectives to conceive the reference unit of spoken language, which can be seen as a unit of *action* (Austin, 1962) or as a unit that expresses a single *focus of conscience* (Chafe, 1994). Also, we noticed that most of the authors consider that prosody plays a fundamental role in the segmentation of the speech into discrete units. For some of them, the reference unit of spoken language coincides with units delimited by terminal prosodic breaks. To others, the reference units coincide with tone units delimited by any kind of break.

Then we analyzed the annotations of the full texts stored in the SLAC resource in order to measure the agreement in respect to the annotation of terminal and non-terminal breaks (Section 4). Also, we calculated a pairwise agreement through seven annotators to measure the consistency on the identification of prosodic breaks despite their theoretical framework (Section 5). Both types of measurements showed that prosodic breaks are theory independent entities, since they are consistently recognized by the annotators. The agreement on terminal breaks identification is higher on dialogic texts than on the monologic ones. This finding matched our expectations, since dialogues are situations in which the interactants exchange communicative acts highly marked by prosody. Conversely, in monologues, the speaker is more focused on the production of a spoken text, and the prosodic boundaries are less perceptually prominent.

Also, the quantitative analysis of the overall and pairwise agreements can be used to understand the nature of the reference unit of spoken language. In our view, the prosodic breaks that identify the reference unit should be those ones that are more salient to annotators, independently from their theoretical perspective. In this respect, the overall agreement analysis showed that in more than 60% of the cases in which an annotator recognized a terminal break, all the other annotators recognized a prosodic break as well (being it terminal or non-terminal). Moreover there is a strong agreement between annotators in detecting a prosodic break of any type (more than 80% of cases). Also, the pairwise agreement measurements have shown that, when an annotator marked a terminal break, the cases in which any other annotator did not mark a break are less than 20% (with the only exception of MRT-CHA that have an agreement on *ANY* measure of 0.77). This data shows that the perceptual identification of terminal breaks is consistent across different theories and perspectives, and for this reason it has to be necessarily considered as one of the starting points for the analysis of spoken language. Hence, in our view, the units delimited by terminal breaks are the best candidates to represent the reference units of spoken language. The good results regarding inter-annotation agreement are also encouraging in the direction of pursuing a more in-depth analysis of the acoustic features that correlate with the perceptual breaks.

Acknowledgements

The authors conceived and discussed togheter all the content of this paper. However, each author gave a specific contribution to the work. In more detail, Alessandro Panunzi ideated the comparison system, coordinated the work, and wrote Sections 1, 4.1, 4.2, and 6; Lorenzo Gregori built the SLAC database, performed the pairwise comparison, calculated the standard agreement, and wrote Sections 3.1, 4.3, and 5; Bruno Rocha developed the Unified Tagset, prepared the data for the SLAC database and for the standard agreement, and wrote Sections 2 and 3.2.

References

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

Barbosa, P. A., & Raso, T. (2018). Spontaneous speech segmentation: Functional and prosodic aspects with applications for automatic segmentation [A segmentação da fala espontânea: Aspectos prosódicos, funcionais e aplicações para a tecnologia]. *Revista de Estudos da Linguagem*, 26(4), 1361–1396.

Blanche-Benveniste, C. (1990). *Le français parlé, études grammaticales*. Paris: Edition du CNRS.

Carletta, J. (1996). Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22, 249–254. Retrieved from <https://www.researchgate.net/publication/220485206_Assessing_Agreement_on_Classification_Tasks_The_Kappa_Statistic>

Chafe, W. (1994). *Discourse, counsciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago, IL: The University of Chicago Press.

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

Davies, M., & Fleiss, J. L. (1982). Measuring agreement for multinomial data. *Biometrics*, 38(4), 1047–1051. https://doi.org/10.2307/2529886

Debaisieux, J.-M. (Ed.). (2013). *Analyses linguistiques sur corpus: Subordination et insubordination en français contemporain*. Paris: Hermès.

Debaisieux, J-M., & Martin, P. (this volume). Syntactic and prosodic segmentation in spoken French. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Den, Y., Koiso, H., Maruyama, T., Maekawa, K., Takanashi, K., Enomoto, M., & Yoshida, N. (2010). Two-level annotation of utterance-units in Japanese dialogs: An empirically emerged scheme. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, & D. Tapias (Eds.), *Proceedings of the 7th language resources and evaluation conference (LREC2010)* (pp. 2103–2110). Valetta, Malta: European Language Resources Association (ELRA).

Deulofeu, H.-J. (2003). L'approche macrosyntaxique en syntaxe : Un nouveau modèle de rasoir d'Occam contre les notions inutiles. *Scolia*, 16, 112–125.

Du Bois, J., Chafe, W., Meyer, C., Thompson, S., Englebretson, R., & Martey, N. (2000–2005). *Santa Barbara corpus of spoken American English*. Philadelphia, PA: Linguistic Data Consortium.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378–382. https://doi.org/10.1037/h0031619

Kibrik, A. A., Korotaev, N. A., & Podlesskaya, V. I. (this volume). Russian spoken discourse: Local structure and prosody. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Kibrik, A. A., Korotaev, N. A., & Podlesskaya, V. I. (this volume). The Moscow approach to local discourse structure: An application to English. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Newbury Park, CA: Sage.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates.

Martin, P. (this volume). Analysis of two examples with the dependency incremental prosodic structure model. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Maruyama, T. (this volume). Segmentation of English texts *Navy* and *Hearts* with SUU and LUU. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Mithun, M. (this volume). Basic units. In S. Izre'el, H. Mello, A. Panunzi, & T. Raso (Eds.), *In search of basic units of spoken language: A corpus-driven approach*. Amsterdam: John Benjamins.

Moneglia, M., & Raso, T. (2014). Notes on Language into Act Theory. In T. Raso & H. Mello (Eds.), *Spoken corpora and linguistic studies* (pp. 468–495). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.61.15mon

Appendix. The Unified Tagset

BREAK TYPE	Chafe and Mithun (CHA, MIT)			Cresti-Raso (CNR)		Izreel (IZR)		
	Symbol	(on DB)	Value	Symbol	Value	Symbol	(on DB)	Value
TERMINAL (T)	.	.	Full fall, terminal fall	//	Terminal break	||	||	Prosodic-set / utterance boun
	?	?	Not specified			|| ↗	|| /	Prosodic-set / utterance bounda (with rising tone)
	!	!	Not specified					
NON TERMINAL (NT)	,	,	Partial fall	/	Non-terminal break	| ↗	| /	Non-terminal Prosodic/ Information Module (PM/IM) boundary with rising tone
	…	…	Not specified			| →	| -	Non-terminal PM/IM boundary with level (or slightly rising or slightly falling) tone
		[0]	Aligned unit without an explicit mark			#	#	Abrupt nonterminal PM/IM ending
INTERRUPTIONS & DISFLUENCIES (+)		-L	(at the end of TU)	[/n]	Interrupted utterance	—	---	Truncated PM/1M
		-	(in the middle ofTU)	[/n]	Retracting (n is the number of retracted words)		[0]	Aligned unit without an explicit mark (Navy, turn 4)
	--	-	(at the end of TU)					
		-	(in the middle ofTU)					
	??		Not specified					
NON TRANSCRIBED (X)		[ntrsc]		[ntrsc]			[ntrsc]	
ALTERNATIVE TAGGING				(*tag*)	Alternative tagging		(*tag*)	Alternative tagging

BREAK TYPE	Kibrik-Korotaev-Podlesskaya (KKP)			Martin (MRT)		Maruyama (MAY)	
	Symbol	(on DB)	Value	Symbol	Value	Symbol	Value
TERMINAL (T)	.	.	Statement	C0	Final conclusive contour, falling and low	/L	LUU with syntactic/interactional cue
	…	…	Incomplete statement, incertitude	Ci	Rising interrogative contour	/R	LUU with a reactive token
	¡	¡	Directive				
	?	?	Question				
	@	@	Vocative				
	!	!	Exclamation (can be combined with other marks, also NT ones)				
NON TERMINAL (NT)	,	,	Default continuity (may be combined with illocutionary force marks)	Cc	Complex contour: slightly falling on stressed syllable, rising on the final syllable. Continuation majeure for long sentences	/sp	SUU with a pause more than 0.1 seconds
	:	:	Used in front of lists, explanations, direct and semi-direct speech	C2	Falling contour, variant of Cc in short or average length sentences	/sd	SUU with a prosodic disjuncture
	---	---	Split (EDU split in 2 parts because another EDU wedges in)	C1	Rising contour		
	==	==	Self-repair in EDUs boundaries	Cn	Flat neutralized contour		
INTERRUPTIONS & DISFLUENCIES (+)		[0]	Aligned unit without an explicit mark (Navy, turn 4)			/F	LUU with a fragment or suspended utterance
NON TRANSCRIBED (X)		[ntrsc]		[ntrsc]		[ntrsc]	
ALTERNATIVE TAGGING	¦	(I)	EDU could be divided into more EDUs				

Index

What is the best way to analyze spontaneous spoken language? In their search for the basic units of spoken language the authors of this volume opt for a corpus-driven approach. They share a strong conviction that prosodic structure is essential for the study of spoken discourse and each bring their own theoretical and practical experience to the table. In the first part of the book they segment spoken material from a range of different languages (Russian, Hebrew, Central Pomo (an indigenous language from California), French, Japanese, Italian, and Brazilian Portuguese). In the second part of the book each author analyzes the same two spoken English samples, but looking at them from different perspectives, using different methods of analysis as reflected in their respective analyses in Part I. This approach allows for common tendencies of segmentation to emerge, both prosodic and segmental.

The comparative work among all the segmentations is stored in the SLAC (Spoken Language Annotation Comparison) database, through which the reader can find all the segmentations compared and analyzed, freely accessible online at https://doi.org/10.1075/scl.94.slac.

The audio files of the examples in the book can be found here: https://doi.org/10.1075/scl.94.audio.

ISBN 978 90 272 0497 4

9 789027 204974

JOHN BENJAMINS PUBLISHING COMPANY