PROSODIC SEGMENTATION AND FUNCTIONAL CORRELATIONS: THE CASE OF JAPANESE

MONEGLIA, Massimo[1]*
CRESTI, Emanuela[2]

[1]University of Florence
[2]University of Florence

_____

Abstract: *This paper presents a pilot study based on the NUCC corpus aimed at verifying the consistency of the Language into Act Theory (L-AcT) for the annotation of information structure in spoken Japanese. L-AcT focusses on the perceptual relevance of prosodic breaks, foresees a strict correspondence between prosodic units and information units and grounds the Information structure on the unit bearing the illocutionary cues (Comment). Although the analyzed data are limited, the pilot confirms the theoretical assumption that the detection of terminal breaks in speech goes hand in hand with the identification of speech acts by competent speaker. The illocutive definition of the Comment is also verified on the basis of pragmatic evidences. The model also foresees a typology of information functions. The main types which pattern the utterance (Topic, Parenthesis, Appendix and Dialogic Units) also fit with the analysis of the Japanese data. The properties of Information structure turn out largely language independent. Japanese word order (SOV) applies within the Information unit, but it does not across information units, as exemplified by post-verbal tails in Appendixes. Beyond the occurrence of morphemes and particles, which usually mark cases and functions in this language, the Topic-Comment Information structure can be performed solely by the prosody. The frequency of information units such as the Topic and the Appendix, instead, seems a language-dependent feature.*

Keywords: *Spontaneous Speech Corpora; Japanese; Prosodic Segmentation; Pragmatics; Information Structure.*
_____

1 Introduction

This paper presents a pilot study aimed at verifying the consistency of the Language into Act Theory (L-AcT) (Cresti, 2000; Moneglia, Raso, 2014; Cresti, Moneglia, 2018) in annotating the information structure of Japanese speech corpora. The pilot is intended to bootstrap the possible development of an annotated spoken Japanese mini-corpus, which will be stored in the IPIC Database (Panunzi, Gregori, 2012).

IPIC is a multilingual collection of spontaneous speech mini-corpora that have been tagged with their information structure according to the L-AcT methodology. Each mini-corpus records a sampling of about 5,000 reference units i.e. utterances and stanzas (see below). Each one complies with the same *corpus design matrix*, allowing cross-linguistic comparisons of information structure properties in the considered languages (Cresti, Moneglia, 2005; Raso, Mello 2012). At present, IPIC has resources for Italian, Brazilian Portuguese, and Spanish (Panunzi, Malvessi-Mittmann; Nicolás-Martinez, Lombán-Somacarrera, 2018) while a comparable mini-corpus of American English has also been delivered by the LEEL lab (Cavalcante, Ramos, 2016; Cavalcante, Raso, Ramos, 2018). The development of a Japanese mini-corpus may represent a significant application of the L-AcT framework for linguistic families outside of the Romance languages and English, helping to validate its information tagging model.

The Japanese dataset relies on the Nagoya University Conversation Corpus, NUCC (Fujimura, et al. 2012), which is one of the largest corpora currently available for spoken Japanese. It is distributed by the National Institute for Japanese Language and Linguistics (NINJAL) and corresponds to approximately 80 hours of conversation and 1.5 million

*Corresponding author: moneglia@unifi.it

transcribed morphemes (Ogiso et al. 2012). The corpus contains 129 natural dialogues and conversations between friends, family members, and colleagues, presenting a large variety of contexts. For this reason, it is a valid source of selection samples which fit with the design model of the IPIC corpora, thus allowing cross-linguistic comparisons in the spontaneous speech domain (Cresti, Fujimura, 2018).

The pilot study considers around 100 excerpts taken from the following recordings:

1. a dialogue between a husband and wife at home, concerning the garden of their house [J090 - garden];
2. a dialogue between two female friends in an office [J018 - chats];
3. a conversation between colleagues in a restaurant [J089 - restaurant];
4. a conversation among students [JL01- after the lecture].[1]

The transcripts are stored in Japanese (Hiragana, Katakana, and Kanji) and have recently been (automatically) transliterated into the Roman characters.[2] The table below gives examples. Specifically, each row gives a word sequence ending with strong punctuation, each one corresponding to an utterance transcribed in terms of standard orthographic criteria. Strong punctuation indicates autonomous propositions, while commas segment them according to major syntactic boundaries and transcriber competence.[3]

Transliteration	Transcript in Japanese characters
M 0 0 3 ： komagoma to soko de sai te n no 、 hiyashinsu ka nee 。	M003：細々とそこで咲いてんの、ヒヤシンスかねえ。
naa 。	なあ。
F 0 1 1 ： sono shita de ？	F011：その下で？
a^a 、 mou de mo nani ka shindoi wa 。	あーあ、もうでも何かしんどいわ。
a 、 to 、 chu^rippu to ka te ima mou ue tara yasui nen kedo ne 、 kyuukon 。	あ、と、チューリップとかて今もう植えたら安いねんけどね、球根。
mou imasara nani ka shindoi wa 。	もう今さら何かしんどいわ。
M 0 0 3 ： chu^rippu nanka ichi tsu mo de te hen yan 、 uchi 。	M003：チューリップなんか１つも出てへんやん、うち。
F 0 1 1 ： iya 、 soya kara atarashiku chu^ 、 kyuukon ue tara 。	F011：いや、そやから新しくチュー、球根植えたら。
ima mou yasuku natsu teru 、 sugoku 。	今もう安くなってる、すごく。
nedan ga 。	値段が。
M 0 0 3 ： dou iu koto 。	M003：どういうこと。
de teru yatsu 。	出てるやつ。
F 0 1 1 ： kau tara yo 。	F011：買うたらよ。
ima kara ueru n yo 。	今から植えるんよ。
M 0 0 3 ： kyuukon kara ？	M003：球根から？
（un） un nai i na 。	（うん）うんないいな。

Figure 1: Transliteration results for the NUCC corpus (Garden excerpt).

The L-AcT methodology envisions the alignment of each utterance (i.e. each pragmatically accomplished speech entity) to its acoustic counterpart (the acoustic segment demarcated by a terminal boundary using the software *WinPitch*), and the annotation of its information structure with respect to a specific tagset (Moneglia, Raso, 2014). L-AcT assumes that an utterance is the counterpart to a speech act and is characterized by an illocutionary accomplishment, tracing back to the pragmatic tradition begun in Austin (1962) and adopted in corpus-based grammars such as Biber et al. (1999). In section 2 of this paper we will briefly detail the main assumptions of the L-AcT model with regard to the prosodic cues necessary for the segmentation of the speech flow into utterances and the utterance into information unit types. In 3 we will challenge the model's criteria for allowing the segmentation of speech flow

---

[1] The acoustic sources of the NUCC transcripts are not available to the public. The copyright owner granted only the wav files specifically for this study.

[2] Examples in this paper cite the exact transliterations delivered in the NUCC corpus.

[3] We will add prosodic segmentation to the original transcripts but will also keep the original punctuation, which frequently - but not always - coincides with the segmentation.

into speech acts on the basis of prosodic and pragmatic cues. In 4 we will verify its consistency with respect to possible internal segmentations of each utterance into information units, as well as the adequacy of the information function tag set used by L-AcT when applied to the Japanese dataset. In 5 we will consider the interface between information structure and syntax and provide support for the adequacy of the model in capturing relevant grammatical properties of the language, especially with regard to *particles* and *word order* (Aoyagi, 2006).

2 The L-AcT model

L-AcT assumes that speech flow may be segmented via pragmatic and prosodic cues into reference units suitable for linguistic analysis. In speech, a reference unit is the highest-ranking unit, "which is autonomous in terms of its pragmatic or communicative function" (Quirk et al., 1985:78)

In this framework a reference unit may belong to one of two types: an *utterance* or a *stanza*, which may or may not contain a verb and do not necessarily correspond to a sentence.

According to speech act theory (Austin, 1962), an utterance is defined as the counterpart to a speech act. From the corpus driven investigation into Romance corpora in Cresti (2018), we see that utterances are characterized by their interactive forces and concern mainly directive illocutions, such as orders, questions, instructions, warnings, introductions, deixis, requests of attention, and so on. They represent the primary reference units for spontaneous speech analysis and 90% of reference units in C-ORAL corpora are of this type.

Conversely, a stanza expresses a flow of thought (going by the definition in Chafe, 1994) and is typical in monologic and professional discourse. It corresponds to a sequence of speech acts that are evaluated within the L-AcT repertory of illocutionary types as weak assertive forces (Cresti, 2010).[4] These speech acts are added by the speaker one after the other, outside of an overall programme, and may continue until the conclusion of the flow of thought. An example of such would be a part of a story or an explanation. We will limit our argument in this paper to utterances only.[5]

In accordance with the tradition (Karcevsky 1931; Crystal 1975; Cruttenden 1997), L-AcT considers that utterances boundaries are demarcated by prosodic breaks that are perceived with the quality of being terminal ('t Hart et al., 1990; Swerts, 1997; Moneglia, Cresti, 2006).

Every utterance is composed of an information pattern which may be simple or complex. Each information unit within an information pattern is performed by a prosodic unit. The prosodic units of a complex pattern are separated from one another by non-terminal breaks. Therefore, in order for the L-AcT model to be applied to a language, two preliminary operations are necessary:

- identification of terminal breaks;
- identification of non-terminal breaks.

In L-AcT's view prosody and information structure belong to independent systems. However, given that prosodic units map one-to-one with information units, the annotation of prosodic breaks is the basis for the identification of information units in the flow of speech. The

---

[4] A description of the L-AcT illocutionary repertory - consisting of about 90 types (Cresti, 2018) - is not the primary goal of this paper, thus no definition or explanation is given for the interpretation of the illocutionary labels.

[5] See Cresti (2010) and Moneglia and Raso (2014) for details on the notion of the *stanza*.

core of the information pattern is one specific information unit known as the Comment, dedicated to the expression of the illocutionary force. The Comment unit is necessary and sufficient for a complete information pattern, since the expression of one illocutionary force specifies how the reference unit should be interpreted. The illocutionary cues are expressed by the Comment unit by means of its prosodic form.

Table 1: The tagset of information functions defined in L-AcT[6]

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.
	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 a single, conventional rhetorical effect.
	Bound Comment	COB	A sequence of weak Comments which are produced by progressive adjunctions following the flow of thought (*Stanza*).
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 stop his uncollaborative behavior.
	Phatic	PHA	Controls the communicative channel, maintaining 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	Works as an emotional support, stressing the sharing of a social relationship.
	Discourse Connector	DCT	Connects different parts of the discourse, indicating their continuation to the addressee.

The information pattern is *simple* if it is composed of just one information unit of the Comment type, it is *complex* otherwise. In complex information patterns, other optional information unit types support the Comment, with each one corresponding to a dedicated

---

[6] Table 1 gives the standard set of information unit tags and their functions as published in Moneglia and Raso (2014) and discussed therein.

prosodic unit and to a specific information function. Information functions are classified into two basic types, depending on whether they work to fulfil the semantic content of the utterance or function in its communicative support (Discourse markers). The list of information unit types along with their tags is found in Table 1.

The aim of the pilot is to verify the adequacy of the L-AcT model for the segmentation of spoken Japanese, according to key operational principles. We will verify in particular the consistency of the *Comment* principle and whether the detection of prosodic breaks allows for the identification of reference units and information units. We will also consider the adequacy of the main information functions defined in L-AcT with respect to the Topic, Appendix, Parenthesis, and Dialogic units. The overall hypothesis that informational relations hold at a cross-linguistic level, independently of language grammar, will be discussed face to the limited dataset provided in the pilot.

3 Terminal breaks, non-terminal breaks and the pragmatic nature of the reference unit

The translation, segmentation into prosodic units, and judgements concerning the autonomy and interpretability of speech segments have been achieved with the assistance of three PhD students in linguistics at Nagoya University. The students have been trained in the recognition of prosodic breaks according to the standard methodology adopted for the processing and validation of the corpora in the C-ORAL-family. The methodology is published in Cresti and Moneglia (2005) and Raso and Mello (2012) and relies on perceptual evidence.

Data on the interrater agreement are available in Danieli et al. (2004), Moneglia et al. (2010), and Raso and Mittmann-Malvessi (2009). Current projects for the automatic detection of prosodic boundaries reach promising results specifically on Brazilian Portuguese speech data (Barbosa 2008; Mittmann-Malvesi, Barbosa, 2016;).

Throughout this pilot study, the prosodic segmentation into terminal and non-terminal prosodic breaks and all judgements concerning the interpretability of speech segments were achieved through consensus agreement among the native speakers. In cases of disagreement, consensus was always reached upon the presentation of the acoustic analysis.

Let's take a look at example 1., extracted from [J090- garden (1-2)].[7] Although 1. constitutes a short, unique turn, two distinct utterances can be identified from the prosodic boundaries, allowing for distinct pragmatic interpretations. In keeping with typical Japanese grammar, the end of the utterances is marked by final particles (in bold in the transcripts).

1. *M3A-1:
細々とそこで咲いてんの
Hosoboso to sokodesaiten**no** //
Secretly there blooming PR //
'Something is secretly blooming over there'
%ill: assertion

---

[7] Within the examples the following information is presented on separate lines: a) the transcription in Japanese characters; b) the syllabic transliteration into Roman characters, with information function tags in the apices; c) the English translation (character by character); d) an overall translation (in square brackets); e) the L-AcT illocutionary classification (21). Speakers are identified by M (male) or F (female), together with a unique id.

*M3A-2:

ヒヤシンスかねえ*?*

Hiyashinsu**ka-nē** ?

hyacinth PR interrogative ?

'(are they) hyacinths?'

%ill: confirmation request

The two utterances are simple from an informational point of view, since they are each composed of a single prosodic unit corresponding to one Comment information unit. The first accomplishes an assertion concerning the blooming of flowers in the garden, while the second is a request of confirmation with regard to the flowers' type. Figure 2 shows that the two utterances are separated by a salient break which is conveyed by a strong F0 reset and a pause. This break is perceivable by non-native speakers, too.
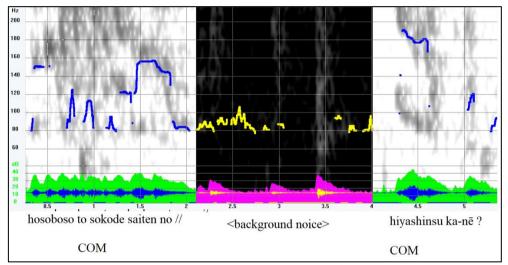

Figure 2: F0 track of example 1.[8]

Although prominent to non-native speakers too, non-natives cannot properly judge the terminal or non-terminal nature of major prosodic breaks. The following two examples presents opposing judgements given by non-native speakers, resulting in interpretations which did not fit with the realities of the speech act performances. The salient break in 2., which is connected to a rising contour, may be perceived by non-native speakers as a continuation, while the salient boundary in 3., which shows a descending contour, is perceived as terminal.[9] Neither utterance is terminated with a final particle.

2. FL01:

十三?

jusan ? COM

thirteenth ? COM

---

[8] The F0 tracks and spectrograms were achieved using the speech software *WinPitch*, which allows an accurate calculation of acoustic parameters for low quality recordings. To ensure the accuracy of the F0 calculation the F0 track is paired with either the first or second formant.

[9] The perceptual judgements by non-natives reported here are not validated. The reader may replicate the author's judgements via the audio files provided.

thirteenth?

FL01:
うち十三…
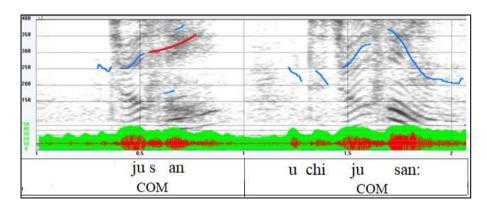uchi jusan: // COM
we thirteenth: //COM
we (are) thirteenth…

Figure 3: F0 track for the first formant in example 2.[10]

3. *M3A18:
もうあんた今ごろ全部, 葉っぱ-が出そろってな-あかんよ。
mou anta imagorozenbu /TOP happa-ga   desorotte   na-akan**yo** //COM
already you now every / TOP leave-SUB   come-out   must   PR FIN // COM
'*As a whole for now leaves had to be already born*'
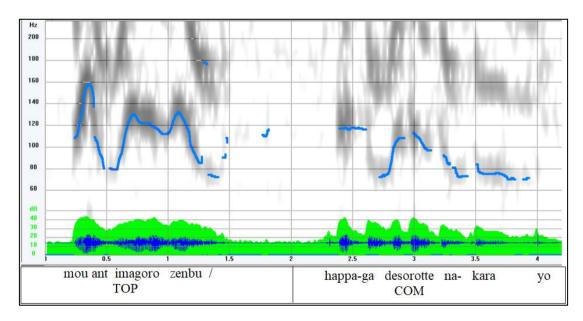%ill: self- conclusion

Figure 4: F0 track of example 3.

---

[10] By positioning the F0 on the first or second formant, calculation errors become more evident. The red line here and below in Figure 7 figures out what the F0 should be like to be more realistic.

As the transcript shows, native speakers easily recognize that the first break in 2. is terminal, since it corresponds to a concluded speech activity (a *request of confirmation*) and is followed by a second speech activity (a *supposition*). If a stretch of speech can be interpreted in isolation as a speech act, the prosodic break is judged to be terminal.

Furthermore, in 3. a native speaker does not assign the value of an independent speech act to the first prosodic unit. The break is perceived as non-terminal since the accomplishment of an illocution cannot be assigned to it in isolation. As a consequence, the prosodic unit is considered part of a sequence that, taken together, is interpreted in terms of the L-AcT repertory as a *self-conclusion*. Therefore, the identification of the terminal quality in a major boundary does not follow from any language independent prosodic properties [rising vs falling boundary tones] but requires strict access to language competence which grounds the pragmatic interpretation of speech. Using this competence, the linguist determines whether the prosodic unit may be interpreted in isolation or not. When it doesn't, the unit is part of a larger utterance and the perceived prosodic break is considered non-terminal.

Thus, the assignment of a terminal or non-terminal value to a perceived prosodic break depends on pragmatic judgment which only native speakers have access to, as predicted by L-AcT. It may be noted that the presence of the final particle [**yo**] in 3. also indicates the end of the utterance, however it is not by any means necessary (e.g. example 2.).

4 The structure of information within the reference unit

4.1 The comment principle

In accordance with the above interpretations, 1. and 2. correspond to a sequence of simple utterances, each one comprised of a single information unit that is concluded by a terminal prosodic break and bearing an independent illocutionary value. Example 2., as well as 4. below, are both good examples for demonstrating that in Japanese, too, illocutionary cues are conveyed specifically by prosody. In both examples, the same locutive content is repeated and no other linguistic index beyond prosody (e.g. final particles) is responsible for the different illocutionary forces assigned to the two utterances in each dialogic turn. In 2., the word 'jusan' [thirteen], performed with a rising contour on the stressed syllable, expresses a *request of confirmation* and the subsequent 'jusan', performed by the same speaker with a lengthened falling contour, expresses a *supposition*.

In 4. the word 'supi^do' [speed], which is performed by the first speaker with a modulated rising contour, is a *request of confirmation*. The response, 'supi^do', performed by the second speaker with a falling contour at a very low F0 level, corresponds to a *confirmation*.

4. F098-18 :
スピード？
supi^do？ COM
speed ? COM
'(*does is depend on*) *speed* ?'
%ill: request of confirmation

F011-19 :
うん、スピード。

supi^do //COM

speed // COM

'(*it depends on*) *speed*'

%ill: confirmation

       This paper is not the place to discuss which specific prosodic parameters correlate with which illocutionary variations, however examples such as 2. and 4. ground the assumption that the prosodic form of the Comment unit is correlated with the performance of speech acts.[11] In a language like Japanese, this role is also played by particles, however, in the absence of particles (as in 5.), the above illocutionary variations may be interpreted only by considering the prosodic performance. The actual interpretation will be totally underdetermined otherwise.
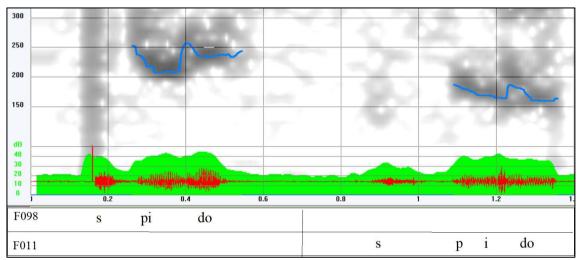

Figure 5: F0 track of example 4.

       The following examples, as well as example 3., allow us to verify the consistency of L-AcT when presented with spoken Japanese, specifically with regard to:

- the segmentation of the utterance into information units, as correlated with the detection of non-terminal breaks;
- the core idea that one specific information unit conveys the illocutionary cues in the utterance

      Moreover, the adequacy of the set of information functions foreseen in the L-AcT model for describing Japanese spoken data is also investigated.

      In 1., 2., and 4., each utterance is comprised of only one prosodic unit ending with a terminal prosodic break. It is considered simple from both a prosodic and informational point of view. Beyond the overall correlation between prosodic performance and speech act variations,

---

[11] The Comment is the information unit dedicated to the expression of the illocution within the utterance. Comments are necessarily performed through a prosodic unit of the type *root*, according to IPO terminology ('t Hart et al., 1990). Root prosodic units record many formal variants whose properties comprise not just F0 variation, but also intensity, duration of syllables, timing, speed, gradation of movements, and accuracy of phonetic execution. A one-to-one correlation between root prosodic types and illocutionary types cannot yet be hypothesized, given the rich repertory of the latter (approx. 90 types). At present, about twenty prosodic forms have been identified conveying distinctive illocutionary values; see Cresti and Moneglia (2018a), Cresti (2018), Cresti et al. (2003), Firenzuoli (2003), and Rocha (2016). Further empirical research is underway.

which is evident in simple utterances, L-AcT foresees that when the utterance is segmented into prosodic units ending with non-terminal prosodic breaks, each prosodic unit constitutes an information unit. This is clear for instance in 3., where the utterance is segmented by both the F0 movements and a pause, with the first prosodic unit corresponding to a Topic unit. What is more interesting in this example, however, is the nature of the second unit; L-AcT assumes that within an utterance characterized by an illocutionary value one and only one prosodic unit identifies the information unit bearing the illocutionary information. This unit is known as the Comment.

This core assumption of the theory may be verified empirically by listening to the units making up a complex utterance in isolation. Only one unit is pragmatically interpretable on its own. In 3. the second unit can, in principle, be interpreted by competent speakers even if the first unit is erased from the acoustic source, whereas the first unit cannot. Let us also consider the dialogue in example 5. between a wife and husband, where the wife complains about a delay in the planting of the tulips and (in 6.) the husband notes that, indeed, nothing flourished.

5. *F1A8:

あとチューリップ とかて今、もう　植え -たら安いねんけどね、球根。

a、to, /[PHA] chu^ripputoka-te /[TOP]　ima, mou /[PAR]　ue -tarayasui-nenkedo**ne**,

/[COM]kyuukon //[APC]

ah well /[PHA]　tulip such-as /[TOP]　　right now /[PAR]　plant-if cheap but PR /[COM]　　bulb

//[APC]

'*ah well, the tulips, if you (had) already planted (them) it would be less costly, the bulbs*'

%ill: expression of disagreement

6. *M3A:

チューリップなんか , 1 つ-も 出てへんやん うち .

chu^rippu　nanka /[TOP]　　hitotsu-mo de-te hen **yan** /[COM]　uchi　//[APC]

tulip such-as /[TOP]　　anything go-out not isn't /[COM]　our place //[APC]

'(*for what regards*) *tulips, nothing flourished, in our place*'
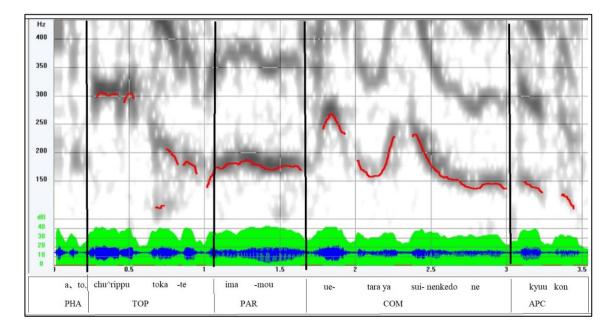
%ill: ascertainment

Figure 6: F0 track for the first formant in example 5.

As the F0 tracks in Figures 6 and 7 show, both of the utterances are segmented into prosodic units by non-terminal breaks and present complex prosodic patterns. The breaks are perceptually quite clear and are marked by F0 resets. Working with competent speakers, we first verified that only one unit plays the role of the Comment and may be interpreted in isolation. In parallel, all of the other units may be erased from the signals without prejudicing the interpretability of the utterances.
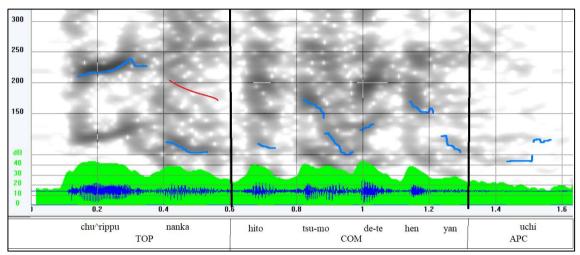

Figure 7: F0 track of example 6.

In other words, the information units tagged as Comments in examples 5. and 6. (given again below in isolation) convey the illocutionary forces of each utterance and receive the pragmatic interpretations of *an expression of disagreement* and *ascertainment*, respectively.

植え -たら安いねんけどね
ue -tara   yasui-nenkedo**ne** /COM
plant-if   cheap   but PR /COM

１つ-も出てへんやん
hitotsu-mo   de-te hen   **yan** /COM
anything   go-out   not   isn't / COM

Therefore, as far as we have seen from the complex utterances in this pilot, the Comment principle seems to hold for Japanese.

4.2 The other information unit types

Beyond the core principle of the Comment, L-AcT's assumption concerning the relation between prosodic parsing and information structure is, in fact, twofold: a) information units within the utterance (identified by non-terminal breaks) play a function at the level of information structure; b) the possible information functions are a closed set that hold at a cross-linguistic level. The Topic-Comment is the basic information pattern, while the Appendix and Parenthesis units constitute supplementary strategies for packaging information. The pilot study

shows that the set of information functions defined in L-AcT can be found in Japanese speech. These are the basic requirements of the functions:

- The information function of the Topic is defined in L-AcT at the pragmatic level. The Topic specifies to the addressee what the illocutionary activity performed by the Comment is about. From a formal point of view, it must precede the Comment and should bear a strong prosodic prominence (*prefix* prosodic form) (Signorini, 2005; Cresti, Moneglia, 2018b; Cavalcante, 2015; Raso, Cavalcante, Mittmann-Malvessi, 2018).). Topic-Comment is a well-formed prosodic pattern.
- The Appendix performs a textual integration of the Comment's content. It has low semantic relevance and behaves as an adjunct at the end of the utterance. It is filled mostly with generic terms, repetitions of previous words, and concluding formulas with the intent of ensuring the addressee's agreement. The Appendix occurs necessarily after the Comment unit and is performed by a prosodic unit of the *suffix* type, with a low-descending F0 profile and weak intensity. It is distinct from the Topic, as it does not specify the domain of relevance of the Comment. Comment-Appendix is a well-formed prosodic pattern.
- For Dialogic units, L-Act foresees that Discourse markers are always isolated from the rest of the utterance by non-terminal prosodic breaks and cannot be interpreted as independent speech acts (Raso, 2014; Raso, Vieira, 2016; Gobbo, 2018; Frosali, 2008; Cresti, 2000; Cresti, Moneglia, 2019).

The above definitions for the L-AcT model match directly with Japanese data concerning the informational role of the Topic and the Appendix, which precede and follow the Comment, respectively. The Topic units found in the Japanese data are coherent with the informational definition given for it in L-AcT. For instance, the *self-conclusion* in 3. is relative to the period of the year; the *disagreement* in 5. and *ascertainment* in 6. concern "tulips". Furthermore, prosodically speaking, Japanese fits in with the general features of the L-AcT model; the Topic bears a strong prosodic prominence while the Appendix is weak and yields a flat F0 movement characterized by a significant decrease in intensity.

The Parenthesis units found in the Japanese data closely follow the properties foreseen for it in L-AcT. For instance, competent speakers verified that the Parenthesis in 5. can be erased without jeopardizing the well-formedness of the Topic-Comment-Appendix prosodic pattern. It's worth noting that the sequence *ima mou* [right now] in 5. was separated by a comma in the Japanese transcript, however the sequence contains neither a pause nor a prosodic reset. The sequence is performed as one prosodic unit that behaves exactly like a Parenthesis information unit. This interpretation has been closely verified with our native-language collaborators, who support the conclusion that *ima mou* is one information unit playing the role of the Parenthesis.

On the contrary, speakers also verified that, if the Topic unit is deleted, the resulting pattern (Parenthesis / Comment / Appendix) does not make sense. This may be due to the prosodic performance, since *ima mou* [right now] might, in principle, be a kind of topical reference for an act of *disagreement*.

The dialogic units (for instance, *a to* [ah well]) in 5. are prosodically isolated. If played in isolation with its actual prosodic form the unit cannot be accepted by competent speakers as an autonomous utterance.

5 Information Structure and Japanese grammar

The systematic annotation of prosodic breaks for the marking of utterance boundaries and the detection of the information functions performed by prosodic units leads to the highlighting of interesting properties at the interface between information structure and Japanese grammar.

First, it is well known that Japanese is a Topic-language (Lee, Tompson, 1976; Shibatani, 1982), and our spontaneous speech pilot confirms this fact. The prefix-root prosodic pattern supporting the Topic-Comment information pattern is indeed very frequent, as is the Topic-Comment-Appendix pattern. It is worth noting that the canonical linear order of information unit types found in Romance and Germanic Languages (Topic-Comment-Appendix) does not vary in Japanese, even though Japanese is a SOV language.

The expressions filling the Appendix unit might apparently contradict the Japanese word order. Post-verbal constituents in Japanese - referred to as *tail*s in the literature (Abe, 2004; Kanada, 2010) - are good candidates for being Appendixes in the L-AcT definition of the term. When considered from a prosodic point of view, *tail*s appear to be performed in a *suffix* type prosodic unit, as is foreseen for Appendixes. According to our interpretation, the element in the Appendix is not an argument in a predicative expression, but functions, syntactically speaking, as an adjunct. Therefore, post-verbal constituents fall outside of the sentence configuration. This is exactly what occurs in the previous examples. For instance, in 5., 球根 [bulbs] might be considered the subject of the predicate; i.e. 'the bulbs should have cost less'. However, this lexical item does not follow Japanese word order, falling in an Appendix unit at the end of the utterance after the predicate in the Comment unit. This is allowed by the information structure which is an independent level with respect to syntax and foresees the Topic-Comment-Appendix language independent order.

For what we can see informally from our pilot, the frequency of the Appendix unit may be higher in this language than in the Romance languages. For instance, the Italian IPIC mini-corpus records only 196 utterances containing Appendixes (3.46% of the total), while most utterances bearing an information structure in our pilot contained an Appendix.

It is also important to stress that in Japanese information functions are conveyed through prosody, beyond the occurrence of morphemes and particles which usually mark cases and functions in this language (Aoyagi, 2006). As a matter of fact, the Topic-Comment structure may also be performed without morphemes or particles (Shimojo, 2006; Nakagawa, 2016) such as with the Topic in 3. and the Comment in 7. below. More specifically, final particles may occur at the end of the Comment unit rather than at the end of the utterance, when the utterance is concluded by an Appendix unit. For example, the final particles in bold in the transcripts of 5. and 6. mark the Comment boundary and not the end of the actual utterance.[12]

Looking more closely at the relationship between information structure and syntax, the L-AcT model draws a sharp distinction between syntactic relations and informational relations. The internal segmentation of a reference unit through prosody gives rise to a set of information units that are considered *islands* from a modal, semantic, and syntactic point of view. From this assumption, it follows that no compositional relations can hold across information units, being bound by informational relations only (Cresti 2014).

---

[12] Lombardi-Vallauri (2014) argues that "Appendixes are Topics", reflecting, in fact, that in Japanese Appendixes may be introduced by "*wa*". A more detailed investigation into the semantic features of Japanese tails seems necessary.

However, L-AcT foresees some restrictions on this overall principle, since the one to one correspondence of "prosodic unit / information unit" is mitigated specifically when a non-terminal prosodic break signals the scanning process of an information unit. Given that the semantic content of an information unit is conceived in its entirety to enact a specific information function (typically a Topic or a Comment), it may turn out to be longer from a syllabic point of view than the "canonical" length of a prosodic unit. In this case, the information unit is segmented by non-terminal prosodic breaks into scanned units. Then, if an information unit is parsed by prosody in this way, the scanning unit does not play an informational role and it is strictly compositional within the information unit that it scans. When scanning occurs in speech, only one part of a scanned information unit bears the perceptually relevant prosodic movement characterizing its informational role. Scanning units do not bear this movement and in Romance languages are always found before the unit bearing the perceptually relevant movement (Cresti, Moneglia, forthcoming).

The above principle has been challenged in this study. When scanning occurs in Japanese, grammatical word order is strictly followed. However, a remarkable difference with Romance data arises, since scanning occurs in Japanese both on the right and on the left. For instance, the Comment unit in 7. bears the relevant prosodic movement on 運転 [unten], but the unit is not autonomous, since according to competent speakers its interpretation strictly requires the predicative particle なの [na no], which occurs in the unstressed unit on the right. Therefore, the predicative particle finds its scope on the *left* according to standard word order rules and the Comment unit is scanned on the right, after the perceptually relevant prosodic movement.

7. F011-21:
あたしとこのね、うちの連れ合いはね、（うん）ものすごーく運転得意な人なの。
atashi to kono ne、/^{TOP} uchi no tsureai ha ne 、/^{APT} (un) nmonosugo^ku**unten** /^{COM}
tokui a hito**na no** 。//^{SCA}
 mine   PR /^{TOP}   my husband PR-Th /^{APT}   hum unbelievable driver / ^{COM}
good person PR PRED//**^{SCA}**
'*my husband (I mean) / is an unbelievably driver / good person is*'
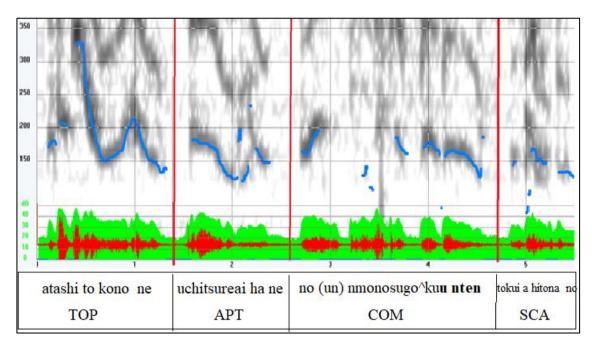%ill: expression of contentment

Figure 8: F0 track of example 7.

The relation between scanning and word order rules may also be verified in the Topic units, which, like the Comment, can be parsed into different prosodic parts. Let's consider the following dialogue in which F011 informs F098 about her son's trip to the U.S.

8. *F011-13:
で、アトランターオーランド間、ディズニーランドまでは、（うん）ええとね、６０
０キロ。

de /INP   atoranta /TOP   (un) o^randokan/SCA   dizuni^ rando made wa /TOP

(un) ee to ne /DCT   ro pyakukiro //COM

well /INP   Atlanta /TOP   Orlando between /SCA   Disneyland until PR-Th /TOP

 (hum) wait /DCT   six hundred kilometers //COM

'well / from Atlanta to Orlando / no, to Disneyland / wait (it's) six hundred kilometers'

%ill: assertion
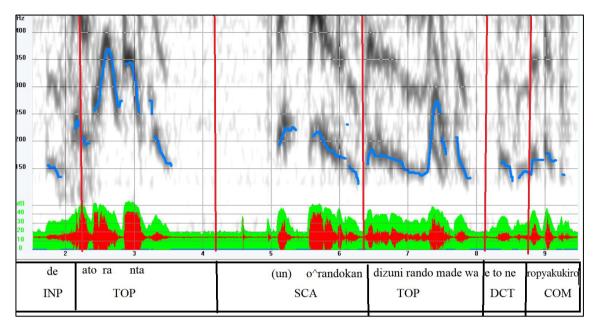
*F098-14:
ふーん。

fu^n //

'hm / hm //'

%ill: back-channel

Figure 9: F0 track of example 8.

9. F011-15：

でも大阪東京間ぐらいやから、それをアメリカやったらすごいあれでしょ。

demo /DCT   Osaka Tokyo kan /SCA   guraiya kara,/TOP   sore wo tabun / TOP
amerika yattara sugoi   /COM   are desho //APC
but / DCT   Osaka Tokyo between /SCA   about is because /TOP   this PR-Obj perhaps /TOP
America by hypothesis very thing/COM   right //APC
'but given that between Osaka and Tokyo / (it is) approximately (the same distance) / this
perhaps / (in) America (is) a very (ordinary) thing / right'
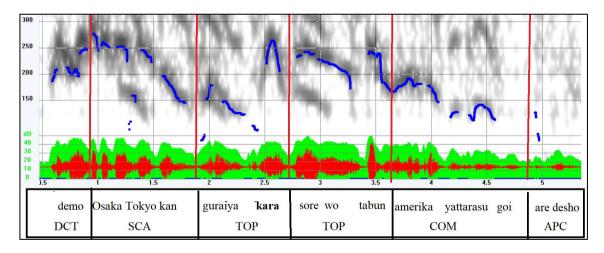%ill: conclusion

Figure 10: F0 track for the first formant in example 9.

The first turn in 8. corresponds to a complex utterance that accomplishes a neutral assertion. Its information pattern opens with a Dialogical unit of type Incipit (Raso, 2014), which is prosodically isolated according to L-AcT. The utterance records a Topic unit (*from*

Atlanta to Orlando), that is scanned by prosody into two parts, and is completed by a second Topic (*until Disneyland*).

Of course, the present pilot is limited and the relationship between canonical word order in Japanese and its order when distributed within the units of the information structure should be studied further. Canonical word order, indeed, seems to hold only within the unit of information. However, we notice, for instance, that the preposition (間 [kan]),[13] which according to Japanese word order must come at the end of the Phrase, links Orlando in the defocused unit on the right compositionally to the unit bearing the Topic movement (*Atlanta*). Also, in this case Japanese presents *scanning on the right* which is compositional with the unit on the left and strictly follows grammatical word order.

A similar phenomenon occurs in the third turn in 9., where the scanning is on the left, as it is usual in Romance languages. The information pattern records, once again, two Topics. The first concerns a comparison of the distance between Tokyo and Osaka and between Atlanta and Orlando which was already presented in the first turn. The first Topic is scanned into two prosodic units, the second of which bears the explicative conjunction から [kara] which hosts the main prosodic movement. The second Topic functions as a modal evaluation.[14] The Comment asserts the fact that a distance like the one between Osaka and Tokyo is common for the U.S. The Comment is concluded by a typical Japanese question tag, with a meaning like '*right*' or '*is not it*'.[15]

In this analysis, the conjunction から [kara] finds its scope in the Propositional Phrase (from Tokyo to Osaka) hosted in a different prosodic unit, however this does not violate the *island constraint* since the Topic is scanned by the defocused part on the left of the main prosodic movement. We have semantic evidence of this analysis. As a matter of fact, neither [o^randokan] in 8. nor ぐらいやから [guraiyakara] in 9. make sense to native speakers without, respectively, the left or the right part of the Topic information unit.

6 Conclusions

The consideration of prosodic performance and, specifically, the identification of terminal and non-terminal prosodic breaks allows the demarcation of utterances and information units in spoken Japanese. The perceptual evaluation of terminal breaks in speech flow is not language independent but goes hand in hand with the identification of speech acts by competent speakers. Information units necessarily correspond to prosodic units, as predicted by many corpus-based studies of information structure (Chafe, 1994). On the basis of the limited amount of data considered in the pilot, L-AcT maps well to Japanese in terms of its basic principles, including for the illocutionary definition of the Comment, which is the information unit allowing the

---

[13] 間 [kan] (roughly "between") is considered a preposition in the standard PoS tagset adopted for Japanese.

[14] The annotation of this unit with the Topic tag might be open for debate. L-ACT foresees the possibility that modals such as "perhaps" may perform a Topic function, since they strongly refer the Comment to the speaker's attitudes and point of view (Cresti, Moneglia 2018b). This case has been tagged in this way while also considering the prosodic prominence of 多分 [tabun] (perhaps), which is coherent with the requirements of the Topic function. Nonetheless, an interpretation as a Parenthesis might also be possible.

[15] Its function fits roughly with an Appendix unit, since it plays the role of a gentle agreement with the addressee.

pragmatic interpretation of an utterance. Japanese is characterized by this central characteristic of information structure, as demonstrated by all utterances tested in the pilot. Beyond the Comment, the main information unit types which pattern the utterance according to L-AcT (Topic, Parenthesis, Appendix, and Dialogic Units), also fit with the Japanese data. The systematic annotation of the correspondence between information structure and prosodic units may also contribute to the grammatical description of the language, particularly with regard to word order and the rules governing particles. Particles seem to mark information units rather than utterance boundaries, but the onset of an information function always correlates with prosodic breaks beyond the presence of particles. At the interface between information structure and grammar, Japanese shows the consistency of L-AcT's island constraint: syntactic compositionality within the information units only. The SOV order does not apply across information units but works fine when prosody scans information units into multiple parts. Likely connected to its word order, Japanese is characterized by the presence of prosodic units which can scan an information unit both on the left and on the right of the unit bearing the functional prosodic movement.

REFERENCES

1. Abe J. On directionality of Movement: A case of Right Japanese Dislocation, *Proceeding of the 58th Conference, The Tohoku English Literacy Society*, 54-61, 2004.

2. Aoyagi H. *Particles and functional categories in Japanese*. Tokyo: Hituzi, 2006.

3. Austin J. *How to do things with words.* London: Arnold, 1962.

4. Barbosa PA. Prominence- and boundary-related acoustic correlations in Brazilian Portuguese read and spontaneous speech. In Barbosa PA, Madureira S, Reis C (eds). *Speech Prosody*. Campinas: ISCA, 2008. pp. 257–260.

5. Biber D, Johansson S, Leech G, Conrad S, Finnegan E. *The Longman Grammar of Spoken and Written English*. London: Longman, 1999.

6. Cavalcante FC. *The topic unit in spontaneous American English*: a corpus-based study. PhD Thesis. Belo Horizonte: Universidade Federal de Minas Gerais, 2015.

7. Cavalcante FA, Ramos AC. The American English spontaneous speech minicorpus. Architecture and comparability. *CHIMERA: Romance Corpora and Linguistic Studies*, v. 3, p. 2, p. 99-124, 2016.

8. Cavalcante FA, Raso T, Ramos A. American English Informationally Tagged Minicorpus. 2018. Available at: www.c-oral-brasil.org.

9. Chafe W. *Discourse, consciousness, and time*: The flow and displacement of conscious experience in speaking and writing. Chicago: UCP, 1994.

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

11. Cresti E. La Stanza: un'unita di costruzione testuale del parlato. In Ferrari, A. (ed). *Sintassi storica e sincronica dell'italiano. Subordinazione, coordinazione e giustapposizione*. Atti del X Congresso SILFI, 713-732. Firenze: Cesati, 2010.

12. Cresti E. Syntactic properties of spontaneous speech in the Language into Act Theory: data on Italian complements and relative clauses. In Raso T, Mello H. (eds.). *Spoken corpora and linguistics studies*, Amsterdam: Benjamin, 2014. pp. 365-410.

13. Cresti E. The empirical foundation of illocutionary classification. In De Meo A, Dovetto FM (eds.). *Proceedings of the International Conference "La Comunicazione Parlata"*. Roma: Aracne, 2018. pp. 243-264.

14. Cresti E, Fujimura I. The information structure of spontaneous spoken Japanese and Italian in comparison: a pilot study. In Manco A (ed.). *Le lingue extra-europee e l'italiano: aspetti didattico-acquisizionali e sociolinguistici*. Milano: Officinaventuno, 2018. pp. 167-189.

15. Cresti E, Moneglia M (eds.). *C-ORAL-ROM. Integrated reference corpora for spoken romance languages*. Amsterdam: Benjamins, 2005.

16. Cresti E, Moneglia M. The illocutionary basis of information structure. In Adamou E, Haude K, Vanhove M (eds.). *Information Structure in Lesser-described Languages*. Amsterdam: Benjamins, 2018a. pp. 359-402.

17. Cresti E, Moneglia M. The definition of the Topic within Language into Act Theory and its identification in spontaneous speech corpora. *Revue Romane*, v. 5, pp. 30-62, 2018b.

18. Cresti E, Moneglia M. The Discourse Connector according to the Language into Act Theory: data from IPIC Italiano. In Bidese E, Casalicchio J, Moroni M (eds.). *La linguistica vista dalle Alpi*. Bern: Peter Lang, 2019.

19. Cresti E, Moneglia M. Some notes on the excerpts according to L-AcT. In Izre'el S, Mello H, Panunzi A, Raso T (eds.). *In search of the basic units for speech*: A corpus-driven cross-linguistic approach to spontaneous spoken communication. Amsterdam: Benjamins, (forthcoming).

20. Cresti E, Moneglia M, Martin P. L'intonation des illocutions naturelles répresentatives: analyse et validation perceptive. In Scarano A (ed.). *Macrosyntaxe et pragmatique: l'analyse linguistique del'oral*. Roma: Bulzoni, 2003. pp. 243-264.

21. Cruttenden A. *Intonation*. Second edition. Cambridge: Cambridge University Press, 1997.

22. Crystal D. *The English Tone of Voice*. London: Edward Arnold, 1975.

23. Danieli M, Garrido JM, Moneglia M, Panizza A, Quazza S, Swerts M. Evaluation of Consensus on the Annotation of Prosodic Breaks in the Romance Corpus of Spontaneous Speech C-ORAL-ROM". In Draxler C, van den Heuvel H, Schiel F (eds.). *LREC 2004: Fourth International Conference on Language Resources and Evaluation*. Lisbon: LREC, 2004. pp. 1513–1516.

24. Firenzuoli V. *Le Forme Intonative di Valore Illocutivo dell'Italiano Parlato*: Analisi Sperimentale di un Corpus di Parlato Spontaneo (LABLITA). PhD Thesis. Firenze: Università di Firenze, 2003.

25. Frosali F. Il Lessico degli ausili dialogici. In Cresti, E. (ed.). *Prospettive nello studio del lessico italiano*, Atti del IX Congresso della Società Internazionale di Linguistica e Filologia Italiana. Firenze: FUP, 2008. pp. 417-424.

26. Fujimura I, Chiba S, Ohso M. Lexical and grammatical features of spoken and written Japanese in contrast: Exploring a lexical profiling approach to comparing spoken and written corpora. In Raso T, Mello H, Pettorino M (eds.). *Proceedings of the International GSCP 2012 Conference: Speech and Corpora*. Firenze: FUP, 2012. pp. 393-398.

27. Gobbo O. *Marcadores discursivos em uma perspectiva informacional*: análise prosódica e estatística. Master Thesis. Belo Horizonte: Universidade Federal de Minas Gerais, 2018.

28. 't Hart J, Collier R, Cohen A. *A Perceptual Study on Intonation*: An Experimental Approach to Speech Melody. Cambridge: CUP, 1990.

29. Kanada K. The Island Effect in Postverbal Constructions in Japanese, *PACLIC* v. 27, pp. 459-466, 2010.

30. Karcevsky S. Sur la phonologie de la phrase. *Travaux du Cercle linguistique de Prague* IV, pp. 188–228, 1931.

31. LEEL. Presents the work of the Laboratório de Estudos Empíricos e Experimentais da Linguagem. Available at: http://www.letras.ufmg.br/leel/.

32. Li C, Thompson S. Subject and Topic: a new typology of language. In Li C (ed.). *Subject and Topic*. New York: Academic Press, 1976. pp. 457- 489.

33. Lombardi-Vallauri E. What can Japanese -wa tell us about the function of Appendixes. *Faits de langue*, v. 43, n. 2, pp. 61-86, 2014.

34. Mittmann-Malvesi M. *O C-ORAL-BRASIL e o estudo da fala informal: um novo olhar sobre o Topico no Portugues Brasileiro*. PhdThesis. Belo Horizonte: Universidade Federal de Minas Gerais, 2012.

35. Mittmann-Malvesi M, Barbosa P. An automatic speech segmentation tool based on multiple acoustic parameters., *CHIMERA*, v. 3, n. 2, pp. 133-147, 2016.

36. Moneglia M, Cresti E. C-ORAL-ROM: Prosodic boundaries for spontaneous speech analysis. In Kawaguchi Y, Zaima S, Takagaki T (eds.). S*poken Language Corpus and Linguistics Informatics.* Amsterdam: Benjamins, 2006. pp. 89-114.

37. Moneglia M, Raso T, Mittmann-Malvessi M, Mello H. Challenging the Perceptual Prominence of Prosodic Breaks in Multilingual Spontaneous Speech Corpora: C-ORAL-ROM/C-ORAL-BRASIL. In *Speech Prosody 2010 – Fifith International Conference*. Chicago, 2010.

38. Moneglia M, Raso T. Notes on the Language into Act Theory. In Raso T, Mello H (eds). *Spoken Corpora and Linguistics Studies*. Amsterdam: Benjamins, 2014. pp. 468-494.

39. Nakagawa N. *Information Structure in Spoken Japanese*: Particles, Word Order, and Intonation, Phd. Dissertation. Kyoto: Graduate School of Human and Environmental Studies of Kyoto University, 2016.

40. Nicolás Martínez C, Lombán Somacarrera M. Mini-Corpus del Español para DB-IPIC. *CHIMERA: Romance Corpora and Linguistic Studies*, v. 5, n. 2, pp. 197-215, 2018.

41. Ogiso T, Komachi M, Den Y, Matsumoto Y. UniDic for Early Middle Japanese: a Dictionary for Morphological Analysis of Classical Japanese. *LREC 2012* Proceedings, pp. 911-915, 2012.

42. Panunzi A, Gregori L.DB-IPIC: An XML Database for the Representation of Information Structure in Spoken Language. In Mello H, Panunzi A, Raso T (eds.). *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation*. Firenze: FUP, 2012. pp. 133-150.

43. Panunzi A, Malvessi-Mittmann M. The IPIC resource and a cross-linguistic analysis of information structure in Italian and Brazilian Portuguese. In Raso T, Mello H (eds.). *Spoken Corpora and Linguistic Studies*. Amsterdam: Bejamins, 2014. pp. 129–151

44. Quirk R, Greenbaum S, Leech G, Svartvik J. *A Comprehensive Grammar of the English Language*. London/New York: Longman, 1985.

45. Raso T. Prosodic Constraints for Discourse Markers. In Raso T, Mello H (eds.). *Spoken corpora and Linguistic Studies*. Amsterdam: Benjamins, 2014. pp. 411-467.

46. Raso T, Cavalcante FA, Mittmann-Malvessi M. Prosodic forms of the Topic information unit in a cross–linguistic perspective: a first survey. In De Meo A, Dovetto F (eds.). *Atti del convegno SLI – GSCP International Conference (Naples 13–15 June 2016)*. 2018. pp. 445-468.

47. Raso T, Mello H (eds.). C-ORAL-BRASIL I: Corpus de referência de português brasileiro faladoinformal. Belo Horizonte: Editora UFMA, 2012.

48. Raso T, Mittmann-Malvessi, M. Validação estatística dos critérios de segmentação da fala espontânea no corpus C-ORAL-BRASIL. *Revista de Estudos da Linguagem*, v. 17, pp. 73-92, 2009.

49. Raso T, Vieira MA. A description of Dialogic Units/Discourse Markers in spontaneous speech corpora based on phonetic parameters. *Chimera*, v. 3, n. 2, pp. 221-249, 2016.

50. Rocha B. Uma metodologia empírica para a identificação e descrição de ilocuções e a sua aplicação para o estudo da Ordem em PB e Italiano. PhD Dissertation. Belo Horizonte: Universidade Federal de Minas Gerais, 2016.

51. Shibatani M. Japanese grammar and universal grammar. *Lingua*, v. 57, pp. 103-123, 1982.

52. Shimojo M. Properties of particle omission revisited. *Toronto Working Papers in Linguistics*, v. 26, pp. 123-140, 2006.

53. Signorini S. Topic e soggetto in corpora di italiano parlato. PhD Thesis. Florence: University of Florence, 2005.

54. Swerts M. Prosodic features at discourse boundaries of different strength. *Journal of the Acoustical Society of America*, v. 101, pp. 514-521, 1997.

55. WINPITCH. Software for prosodic research, with on the fly aligner, real-time spectrograph, multi-tracking F0 analysis, video and audio analysis, and much more (Free installation password required after 30 days of use). Available at: https://www.winpitch.com/.