Information Structure-Prosody Interface: 
Towards a Model of Albanian Intonational Phonology

Enkeleida Kapia\(^1\), Alejna Brugos\(^2\)

\(^1\) Center for Albanian Studies, Albania \(^2\) Boston University, United States

ekapia@bu.edu, abrugos@bu.edu

Abstract

This paper examines the prosodic realization of four information structure correlates in Albanian, i.e., topic, contrastive topic, rhyme, and contrastive rhyme. Nine participants were recruited for a reading task in which sentences were manipulated with regard to their information structure via preceding questions. The results demonstrate that speakers systematically mark these four distinct syntactic constructs with differing prosodic patterns. The results from this data set suggest that the categories of contrastive rhyme and contrastive topics are indeed a distinct categories from the categories of rhyme and topic, respectively. Further, this paper offers a preliminary analysis of tone patterns of Albanian intonation in the ToBI framework.

Index Terms: information structure, Albanian intonation, Auto-segmental Metrical model, ToBI

1. Introduction

Information structure is the packaging of a sentence which explains how new information is added to a specific discourse context (Firbas, 1964 [1], 1971 [2]; Chafe, 1976 [3] etc.). According to the information structure literature, utterances may contain something already known to the discourse context, labelled as topic, and something entirely new, labelled as rhyme, which is usually the information added to the discourse (for a review of constructs and terminology, see Kruijff, 2001 [4]).

A more recent tradition of research has characterized and classified the information structure divisions within a sentence into two layers, that of topic and rhyme, and the layer of contrast (e.g., Vallduví & Vilkuna, 1998 [5]; Vallduví & Engdahl, 1996 [6]; Vilkuna, 1994 [7]; Steedman, 2000 [8]), which enables rhyme and topic to be either contrastive or non-contrastive. For clarity, these four categories are represented in the English examples (1) – (4) below, together with the contexts that trigger them. In (1) and (2), the proper name “Beni” is an already known piece of information, once a (non-contrastive) topic in (1) and once a contrastive topic in (2). In (3) and (4), the proper name “Beni” is a new piece of information, once a (non-contrastive) rhyme in (3) and contrastive rhyme in (4).

The theory of information structure for Albanian has traditionally recognized two information structure concepts, i.e., topic (also labelled as T here) and rhyme (also labelled as R here) (Barri, N., 1975 [9]; Floqi, 1969 [10], 1976 [11]; Demiraj, 1975 [12]; Totoni, M., 1971 [13], 1981 [14]; Rrushi, 1984 [15], Pernaska, R., 1976 [16], 1986 [17]). Recently, however, Kapia (2014) [18] has brought forth new syntactic evidence of the existence of both contrastive rhyme (also labelled as R\(_k\) here) and contrastive topic (also labelled as T\(_k\) here), as in the English examples (2) and (4) listed below.

\begin{enumerate}
\item Q: What did Beni eat?  
A: \{Beni \textit{topic}\} ate meatballs.
\item Q: What did Beni and Roza eat?  
A: \{Beni \textit{contrastive topic}\} ate meatballs.
\item Q: What happened?  
A: \{Beni \textit{rheme}\} ate meatballs!
\item Q: Who ate meatballs?  
A: \{Beni \textit{contrastive rheme}\} ate meatballs.
\end{enumerate}

The aims of this paper are twofold: a) to bring forth a first attempt at the prosodic correlates of these four distinct information structure constructs in Albanian, and b) to provide a preliminary model of intonational phonology in the language using the ToBI framework.

2. Background

2.1. Information structure and prosody interface

Most previous research on the prosodic realization of information structure constructs in other languages focuses on the realization of rhyme, contrastive rhyme, topic, and contrastive topic (Rossi et al, 1981 [19], 1999 [20]; Touati, 1987 [21]; Cristo, 1998 [22]; Le Gac, 2000 [23]; Baltazani & Jun [24], 1999; Alcoba & Murillo, 1998 [25], etc.). Beckman and Pierrehumbert (1986) [26] and Steedman (1991) [27], among other authors, suggest that intonation is directly correlated with information structure. Beckman and Pierrehumbert (1986) identified six pitch accents in English and classified them as topic-rheme markers, which are \{L-H*, L*-H\} for topic, and \{H*, L*, H+L, H+L*\} for rhyme. As far as pitch accents are concerned, Beckman and Pierrehumbert (1986) suggest that the characteristic bitonals are L*-H for topic and H+L* for rhyme. Additionally, L+H* has been reported as a marker of contrastive topics, whereas the pitch accent H* as a marker of contrastive focus (Pierrehumbert and Hirschberg, 1990) [28].

2.2. Prosodic studies of Albanian

Research on Albanian phonology and intonation is quite sparse with the exception of a few studies (Memushaj, 2009 [29]; Jubani 2012 [30]; Boriçi, 1987 [1987]; Themistocleous
& Müller, 2015 [32]). None of these, however, have looked at the interface between information structure and prosody. The current study aims at filling the existing gap through a model of intonational phonology.

2.3. ToBI and IPrA

The ToBI framework (for Tones and Break Indices) is a system for transcribing the phonological categories of the intonation of a language (Beckman et al, 2005) [33]. It includes conventions for labelling tonal patterns (the Tones) associated with prominences (pitch accents) and phrase boundaries (edge tones), as well as labels for indicating the perceived level of disjunction between words (the Break Indices). These labels are typically applied in tiers (such as Praat TextGrids) that are time aligned to the recorded soundfile. ToBI conventions also include tiers for labels for words, as well as a “miscellaneous” tier, which can be used to indicate other events that affect the speech signal, but that are not straightforwardly prosodic, such as coughs, laughter or recording interference.

ToBI tone labels do not incorporate a direct record of acoustic measures, but rather categorizations of prosodic aspects of the signal that can be generalized across speakers and speech contexts. Tone events in ToBI are considered to be targets, or sequences of targets, consisting of a constrained set of tonal levels, primarily H (High) and L (Low). These tonal targets combine in a range of ways, with different alignments with respect to prominent syllables and the segmental string, reflecting their role as markers of prominence or phrase edge. These relatively sparse labels can be used to describe the varying f0 contours of utterances in ways that foster the comparison of exemplars across speakers and contexts.

The ToBI framework was originally developed for the study of the intonation of Mainstream American English (now called MAE, ToBI, Beckman et al 2005 [33]). Methods of the framework have since been applied to the analysis of the prosodic systems of a large number of languages (See Jun, 2004 [34], 2014 [35] for an overview). The development of a ToBI model of a language requires extensive research into the prosody of that language, in order to determine its phonological categories (Beckman et al, 2005) [33]. However, established labels from the ToBI framework can and have been used as tools for generating and evaluating hypotheses for a language’s prosodic categories (Jun & Fletcher, 2014) [36]. In that spirit, Prieto, Jun, & Hualde (2015) [37, 38] have proposed the IPrA (International Prosodic Alphabet), whereby ToBI-type labels can be used to produce a broad phonetic transcription of the tonal patterns as the researcher forms hypotheses about the categories. With the proposed IPrA labelling conventions, the “categorical nature of the prosodic system of a language/dialect” can be transcribed “before distinctiveness of a prosodic feature is known to the researcher.” (Hualde & Prieto, 2015) [39]. Proposed IPrA labels consist of elements from established ToBI conventions, but used more richly, reflecting broad phonetic transcription, until the phonetic variation associated with a hypothesized tonal category has been better understood.

3. Materials and Method

The data used for this study consisted of read speech. There were four types of target sentences with four sentences per type counterbalanced with four filler items. Each sentence had an SVO order and contained one target word. The target word was a disyllabic proper name. Four male and five female subjects (one of which was the first author) participated in the experiment. Participants were monolinguals (with the exception of two) born in Tirana and had no speaking or hearing disability. Subjects read question-answer pairs (exemplified in (1-4) above), which were presented on a computer screen and recorded in a language acquisition laboratory. They were explicitly told to produce the answers according to the questions. The interviewer was a native Albanian speaker who was trained to ask the participant to repeat the question-answer pairs in cases of discrepancies concerning the understanding of the task. Six subjects read each experimental item (4 context types x 4 target sentences) only once, and the subsequent 3 subjects were recorded reading all experimental items a second time for a total of 192 recorded tokens. As a first stage of analysis, a native speaker of Albanian (the first author) listened to all examples to determine with each was a felicitous production in response to the given context question. This judgment was made based on auditory presentation alone. 16 tokens (10 from the T condition, 6 from the Rh_k condition) were excluded from analyses as being inelicentious productions. The remaining 174 experimental items were prosodically labelled.

Soundfiles were analyzed in Praat (Boersma & Weenink, [40]). Items were labeled for target context and words. Tonal event for pitch accents and one level of boundary were annotated. The ToBI conventions state that a full ToBI transcription should include 4 tiers of labels: tones, words, breaks and miscellaneous. However, in the interest of space, only the Tone labels will be discussed in this paper.

4. Results

The model for the description of Albanian information structure-prosody interface follows the ToBI framework (Beckman et al, 2005) and the proposed IPrA conventions (Hualde & Prieto, 2015).

Figures 1 through 4 show examples of responses to the four types of context questions, with differing pitch patterns for each. Figure 1 shows a response to the question, Çfarë bleu Roza? (“What did Roza buy?”). In this example, the subject (Roza) is the topic. This example shows the pitch pattern of a Low on the accented syllable, followed by a rising pitch to a perceived phrase edge, a sequence labelled as L* H%.

![Figure 1: The phrase Roza bleu një libër (“Roza bought a book”) as a response to “What did Roza buy?” Here, Roza is the topic.](image)

Figure 2 shows the phrase Beni hëngë gjofte (“Beni ate meatballs”) as a response to the question Çfarë hëngënë Beni dhe Roza? (“What did Beni and Roza eat?”). In this example, the subject (Beni) is a contrastive topic. The subject noun is
realized with a H* pitch accent, a high but without either a steep initial rise, or a sharp following fall.

Figure 2: The phrase Beni hëngri qoftë (“Beni ate meatballs) as response to “What did Beni and Roza eat?” Here, Beni is the contrastive topic.

Figure 3 shows the phrase Beni hëngri qoftë (“Beni ate meatballs) as response to the question Çfarë ndodhi? (“What happened?” Here Beni is the theme, and is marked by high f0 across the accented syllable, and both a sharp initial rise and a sharp following fall. This pitch pattern is labelled L+H*+L. The phrase also includes a second pitch accent, this time labelled L+H*, on the object noun, and ends in a Low edge tone, labelled L%.

Figure 3: The phrase Beni hëngri qoftë (“Beni ate meatballs) as response to “What happened?” Here Beni is the contrastive rheme.

Figure 4 shows the phrase Roza bleu një libër (“Roza bought a book”) as response to the question Kush bleu një libër? (“Who bought a book?”). Here Roza is the contrastive rhyme. The subject is again marked by a pitch pattern labelled L+H*+L. However, in this case, there is no other pitch movement in the phrase: the object noun does not appear to bear a pitch accent. Instead, the f0 stays low and level to the end of the phrase, a sustained level Low edge tone labelled L-%

Figure 4: The phrase Roza bleu një libër (“Roza bought a book”) as a response to “Who bought a book?” Here Roza is the contrastive rheme.

Below is an inventory of the labels that were used in this data set. These include 9 candidate pitch accent types, annotated with the following symbols:

- **H**: The prominent syllable is marked by relatively high pitch, but without a local steep rise or fall
- **L**: The prominent syllable is marked by relatively low pitch, but without a local steep fall or rise
- **L+H**: The prominent syllable is marked by relatively low pitch, followed by a rise to higher pitch on or around the following weak syllable
- **L+H**: The prominent syllable is marked by a steep rise near or before the syllable onset, with high pitch reached during the prominent syllable
- **L+H+L**: The prominent syllable is marked by a steep rise near the syllable onset, with high pitch reached during the prominent syllable, and then followed by a sharp fall in the following weak syllable.
- **H+L**: The prominent syllable is marked by relatively high pitch, but without an initial steep rise, and then followed by a sharp fall in the following weak syllable.
- **!H**: The prominent syllable is marked by pitch in a mid-range, i.e. relatively high pitch in a locally compressed pitch range. The pitch lower than that preceding syllables bearing a H tone pitch accent, but higher than the bottom of the speaker’s pitch range.
- **H+L**: The prominent syllable is marked by a local fall from high pitch on a preceding weak syllable.
- **?**: A syllable bears some cues to prominence, but with ambiguous or weak pitch cues.

The data set also included 4 edge tone patterns that were labelled with the following symbols:

- **L%**: A final fall to the bottom of the speaker’s pitch range at the end of a phrase
- **L-%**: A sustained low level f0 following a fall to low pitch earlier in the phrase, and continuing to the end the phrase
- **H%**: A phrase final rise
- **H-%**: A phrase final sustained high or mid-range level f0, and continuing to the end the phrase

These symbols represent broad phonetic transcription of the tonal events, and should be considered a “first pass” at a phonological system. For example, the tri-tonal label of L+H*+L was used to indicate pitch accents with a sharp rise and fall, even though further analyses may indicate that the L tone is a retracted edge tone, such as has been analyzed for English (e.g. Barnes et al, 2010) [41]. It is also likely that some pitch accent types listed may be phonetic variants of the same category (e.g. items labelled H*+L and H* may well reflect realizations of the same phonological category.) Likewise, it is possible, indeed likely, that this inventory of tone labels is incomplete, as data were collected in rather limited contexts.

Only one level of phrasal boundary (boundaries above the level) was marked in this data. However, additional research (including measures of duration of preboundary lengthening and pauses in production data, as well as perception data) are expected to reveal additional levels of phrasing, such as distinctions between intermediate and full intonational phrases.
It may also be that subtle, but perceptually salient, differences in segmental alignment are not captured in these labels. Additional experimental work will be necessary. Analyzing pitch accent contours using a quantitative measure such as TCoG (Tonal Center of Gravity) (Barnes et al., 2012) [42] may prove fruitful for characterizing differences in alignment.

5. Discussion

Tables 1 and 2 show the distribution of pitch accent type by elicited context for subject and object.

Table 1. Subject noun pitch patterns, and percentages of their occurrence in the different information status contexts: theme (T), contrastive theme (T_k), rheme (R), and contrastive rheme (R_k).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T</td>
<td>T_k</td>
<td>R</td>
<td>R_k</td>
</tr>
<tr>
<td>L*</td>
<td>3%</td>
<td>2%</td>
<td>6%</td>
<td>0</td>
</tr>
<tr>
<td>L* H%</td>
<td>37%</td>
<td>19%</td>
<td>23%</td>
<td>0</td>
</tr>
<tr>
<td>L*+H</td>
<td>50%</td>
<td>27%</td>
<td>36%</td>
<td>12%</td>
</tr>
<tr>
<td>Percentage L*-patterns</td>
<td>90</td>
<td>48</td>
<td>65</td>
<td>12</td>
</tr>
<tr>
<td>H*</td>
<td>10%</td>
<td>29%</td>
<td>21%</td>
<td>10%</td>
</tr>
<tr>
<td>L*+H*</td>
<td>0%</td>
<td>23%</td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>L*+H+L</td>
<td>0%</td>
<td>0%</td>
<td>4%</td>
<td>71%</td>
</tr>
<tr>
<td>H*+L</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>2%</td>
</tr>
<tr>
<td>Percentage H*-patterns</td>
<td>10</td>
<td>52</td>
<td>35</td>
<td>88</td>
</tr>
</tbody>
</table>

(Note: these numbers collapse L+H* L% and L+H* into L+H*+L, 1 token of each.)

Table 2. Object noun pitch patterns, and percentages of their occurrence in the different information status contexts: theme (T), contrastive theme (T_k), rheme (R), and contrastive rheme (R_k).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T</td>
<td>T_k</td>
<td>R</td>
<td>R_k</td>
</tr>
<tr>
<td>L*</td>
<td>0%</td>
<td>19%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>L*+H</td>
<td>0%</td>
<td>0%</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>Percentage L*-patterns</td>
<td>0</td>
<td>19</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>H*</td>
<td>29%</td>
<td>29%</td>
<td>27%</td>
<td>0%</td>
</tr>
<tr>
<td>L*+H*</td>
<td>24%</td>
<td>8%</td>
<td>38%</td>
<td>0%</td>
</tr>
<tr>
<td>L*+H+L</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Percentage H*-patterns</td>
<td>53</td>
<td>37</td>
<td>65</td>
<td>0</td>
</tr>
<tr>
<td>H+!H*</td>
<td>42%</td>
<td>29%</td>
<td>23%</td>
<td>2%</td>
</tr>
<tr>
<td>!H*</td>
<td>5%</td>
<td>15%</td>
<td>10%</td>
<td>29%</td>
</tr>
<tr>
<td>No accent</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>69%</td>
</tr>
<tr>
<td>Other patterns</td>
<td>47</td>
<td>44</td>
<td>33</td>
<td>100</td>
</tr>
</tbody>
</table>

(Note: for R_k: 6 tokens of the !H* were labelled as *!, but tonal patterns were compatible with !H*. Note also that for T_k, L*-marked object nouns are followed by a H% boundary.)

First, this data set shows that speakers systematically provide prosodic cues related to the information structure constructs of Albanian. While there is no one-to-one mapping between context/construct and tonal patterns, speakers show strong preferences and disparities for pitch accent type in the different conditions. Specifically, the data shows that speakers prefer different tonal patterns for the different conditions represented here similar to the Beckman and Pierrehumbert study (1986) [26], i.e., topic is marked with L* H%, L*+H, contrastive topic is marked with H*, L* H%, L*+H, rheme is marked with L*+H, L* H%, H* and contrastive rheme is marked with L+H*+L. These differences are even more accentuated if the analysis takes into account, not just the target word itself in the subject position, but the object noun located at the end of the same phrase. This suggests that the global prosodic pattern of the phrase contributes to the marking of information structure, in addition to the identity of individual pitch accents.

Secondly, this study brings forth evidence that supports the thesis for human languages in general of Vallduvi and Vilkuna (1998) [5], and of Kapia (2014) [18] for Albanian, that contrastive rheme is a distinct category from all the other information structure categories. Indeed, the data from this study show that contrastive rheme is one of the conditions that is sharply different from all the other three with 88% of its subject noun pitch accents being H* patterns, while the constructs of topic, contrastive topic, and rheme are marked largely via L* patterns.

Thirdly, this study demonstrates a clear difference between the construct of topic and contrastive topic with topic being marked largely via L* tonal patterns, while contrastive topic being marked with patterns that represent its status as topic and patterns that represent status as contrast, i.e., with L* patterns linked to topic and H* patterns linked to contrast. This finding also aligns well with work from syntax which suggests that these two notions are distinct categories (e.g., Vallduvi and Vilkuna, 1998 [5]).

6. Conclusions

This study offers the first contribution towards a general description of the tonal structure in Albanian. In this experiment naïve listeners (with the exception of the first author) engaged in a task where the information structure of SVO sentences was manipulated via preceding questions. Specifically, this study investigated the prosodic realization of four different information structure categories evidenced in Albanian, i.e. topic, contrastive topic, rheme, and contrastive rheme. Results support the prosodic marking of these four distinct constructs in Albanian.

While results of this study are promising, this paper marks only one step into the investigation of Albanian intonation, a subject for which published research is exceedingly sparse. As there were limited contexts and a small set of speakers for the current study, a natural question that arises is to what extent the current results can be generalized. In order to answer this, further studies must take into account a broader data set, with more subjects, and more tokens. We plan to undertake additional production studies to expand on these results, and to further explore the acoustic correlates (including duration measures and more detailed alignment of tones with the segmental string) of the proposed prosodic categories. It is also beyond doubt that perception experiments are necessary to verify the tonal analyses proposed in this paper.
7. References


