



Prosodic and non-prosodic cues to prominence and boundaries: evidence from an RPT study in Albanian

Alejna Brugos¹, Enkeleida Kapia²

¹Simmons University, Boston

²Institute for Phonetics & Speech Processing, Munich & ASA, Tiranë

abrugos@bu.edu, ekapia@bu.edu

Abstract

This paper reports evidence from the Rapid Prosody Transcription method (RPT, cf. Cole & Shattuck-Hufnagel, 2016) in the study of a little-described language. We bring forth results from two perception experiments on the prosody of Albanian, which provide important evidence in the perceptibility of prosodic and non-prosodic correlates in the language's prosodic system, shedding more light on its complexity. Albanian listeners in this study show moderate to substantial agreement in the perception of both prosodic prominences and boundaries, conforming to what has been found in RPT studies from other languages so far (Cole et al, 2010; Baumann & Winter, 2015; Riesberg et al, 2020). In doing so, listeners' perception of prosodic structure correlates with both prosodic cues, i.e. PoLaR intonational categories (Ahn et al, 2019), and non-prosodic cues, i.e. syntactic break and part of speech, suggesting that these factors may serve as cues to prominence and boundary perception.

Index Terms: Rapid Prosody Transcription, Albanian, prosodic prominence, prosodic boundaries, non-prosodic cues

1. Introduction

This paper is concerned with Albanian spoken in the Republic of Albania, located in southeastern Europe. Albanian consists of two main dialects, i.e. Gheg and Tosk, and a standard variety, institutionalized in a state mandated political congress in 1972 (Gjinari, 1988; Kostallari, 1984). It is precisely this variety that will be the focus of investigation in this study. As outlined elsewhere (Kapia et al, forthcoming), empirical phonetic investigations belong to a young tradition in Albanology, among which, investigations of Albanian prosody are even younger.

The only and earliest attempt, intuitive in nature, that describes the prosodic system of Albanian maintained that phonological resources in Albanian, such as f_0 and pauses, expound its grammatical meaning (Boriçi, 1987). More recent empirical work in the ToBI theoretical framework of analysis has suggested that prominence in Albanian is marked by both the head and the edge of the phrase (Kapia et al, 2020; Kapia et al, forthcoming).

2. Methods and Analysis

Given the emerging evidence about Albanian being a head-and-edge prosodic type language, the purpose of this study was to address the question of how native speakers of

Albanian interpret prosodic and non-prosodic cues if required to judge the presence or absence of prominences and boundaries. We therefore conducted two perception experiments using the Rapid Prosody Transcription (RPT) method (Cole et al. 2010a; Cole et al. 2010b, Cole & Shattuck-Hufnagel 2016:7–13). In the RPT setup, naïve ordinary listeners are given minimal instructions in an audio-recordings listening task, while they see on the screen a printed transcript of the recorded excerpts, in which punctuation and capitalization have been removed. Participants are asked to click on words which they perceive as prominent (prominence experiment), or to click on the word which they perceive to be the last word of a grouping (boundary experiment). The advantage of this method is its simplicity and directness; also, elicited this way, prominence and boundary judgments may reflect not only prosodic factors, but also morpho-syntactic, semantic and pragmatic factors. Our main concern here is to examine which prosodic and non-prosodic factors may play a role in listeners' perceptual judgment.

2.1. Subjects and stimuli

The raters of both perception studies were the same 26 native speakers of Albanian, (13 F, mean age 43.7). Participants were persons with no previous experience in prosodic analysis. None reported any hearing or language problems.

The participants annotated the same 20 excerpts of audio recordings for both the prominence and boundary experiments. These excerpts, from 2 male and 2 female speakers, taken from a larger corpus of natural speech, described 2-picture sequences from the Questionnaire on Information Structure (Skopeteas et al, 2006). Excerpts thus consisted of monologues and varied in length from approximately 3 to 15 seconds.

2.2. Test Variables & Data Annotation

We investigated the influence on the native listeners' judgments of prominences and boundaries of a number of prosodic and morpho-syntactic correlates, which have been found to have an effect on prominence and/or boundary perception in other languages, predominantly of Indo-European origin (Riesberg et al, 2020, Bauman & Winter, 2018). For each test word in both experiments, we tested the role of the following three factors: whether listeners' judgments were correlated with our prominence and boundary PoLaR style analysis labels (Ahn et al, 2019), as well as with the factors of 'syntactic break' and 'part of speech' (with a future aim to investigate more prosodic and non-prosodic factors).

As a step in our exploration of prosodic cues, we labeled the sound files using PoLaR, a phonetically transparent prosodic annotation system that unbundles the tagging of prosodic structure from the characteristics of the F0 signal that cues this structure. A native speaker of Albanian (the second author) provided PoLaR annotation for three levels of phrasal prominence (no prominence/prom_0, strong prominence/prom_1, and extra strong/prom_2) and four levels of boundary strength (with the lowest being no phrasal boundary). We chose PoLaR annotation as a flexible framework to facilitate the exploration of f0 characteristics of hypothesized prosodic categories. Additional PoLaR annotations for F0 aspects of the signal were also annotated, including details on F0 range, and the location and scaled levels of turning points associated with prosodic structure. These and other acoustic cues to prosodic structure (such as those from duration and voice quality) will be explored in future work. The F0 and spectrogram of a sample utterance, along with PoLaR annotations for prosodic structure and F0 ranges, are shown in Figure 1.

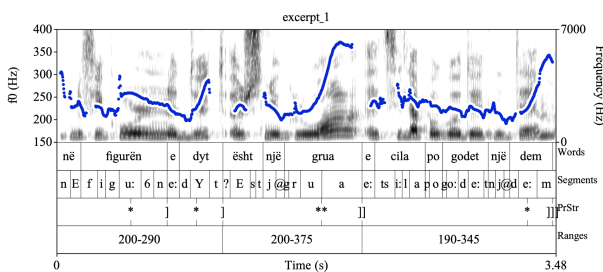


Figure 1: A sample showing partial PoLaR annotations for the excerpt *në figurën e dytë është një grua e cila po godet një dem* ‘in the second picture, there is a woman who is hitting a bull’. In the PrStr tier (for prosodic structure), prominence is marked by 0, 1 or 2 stars, with 0 indicating no perceived phrasal prominence. Boundary strength is marked by 0 to 3 brackets, with 0 indicating no perceived phrasal boundary.

Utterances were also coded for syntactic structure. Relevant to Albanian sentence structure, the label weak syntactic break was assigned to sentence-medial words that were followed by a subordinate clause (e.g. relative clause or prepositional clause), while the label strong syntactic break break was assigned to sentence-final words. As mentioned above, these structural factors were chosen based on what has been suggested about other Indo-European languages (Riesberg et al, 2020; Bauman & Winter, 2018), since little is known about these factors and their relation to prosody in Albanian. For instance, in West Germanic languages, parts of speech that are function words are usually less prominent than content words (Büring 2012; Roy et al, 2017). Syntactic break cues have also been reported to play a role in German as well as in Papua Malayan (Riesberg et al, 2020).

2.3. Data Analysis

Both experiments consisted of binary classification tasks. In the prominence experiment (Experiment I), participants’ marked prominent word choices were coded as 1, while all else was given a 0 code. Similarly, in the boundary experiment (Experiment II), there was a choice for each word to be

marked as either at the end of a grouping or not; these marked words were coded as 1, the rest as 0. Given that our set of 20 excerpts consisted of 384 words altogether, each participant thus produced 384 data points in both the prominence and the boundary experiment. A one-way ANOVA was used to compare the effect of the independent variables on the dependent variables (p-score and b-score) in R (R Core Team 2015).

Subsequently, we calculated the Fleiss’ kappa coefficient, which provides a single coefficient as a measure of agreement across all raters. In addition, we calculated the prominence-score (p-score) and the boundary-score (b-score), which serve as relative measures representing the ratio of subjects that clicked on a word, i.e. that perceived a word as prominent, or that perceived a prosodic break, with respect to the total number of participants. Figure 2 shows the p- and b-scores of the sentence in Figure 1.

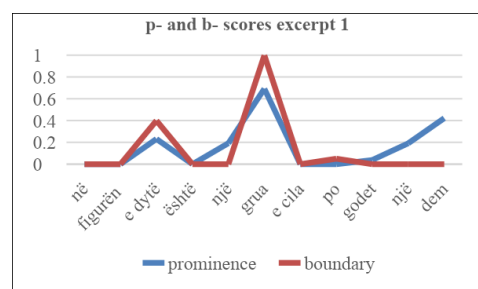


Figure 2: P- and b-scores for the excerpt in Fig. 1. The higher the value, the more participants perceived a word as prominent (blue line) or perceived a boundary after the respective word (red line). Scores here are shown as a proportion of all 26 listener judgments.

3. Results

3.1. Prominence Experiment Results

The Fleiss’ kappa score for the prominence experiment amounts to 0.32. Even though it appears to be a low score, it still is compatible with some other languages, such as German (Riesberg et al, 2020) and English (Cole et al, 2010). In Table 1, we compare the Albanian inter-rater agreement scores from the prominence experiment with those of other studies on American English (Cole et al. 2010), German (Baumann & Winter 2015) and Papua Malayan (Riesberg et al, 2020). Albanian is more comparable with the German and the English scores than the Papua Malayan one, with which there is a difference of a greater magnitude, and for which it has been hypothesized that there is no systematic prominence marking (Riesberg et al, 2020).

Table 1: Fleiss’ kappa scores for prominences in German, English, Papua Malay and Albanian

	German	English	PM	Albanian
Fleiss’ Kappa	0.53	0.42	0.103	0.32

When examining the role of ‘syntactic break’, as seen in Figure 3, a one-way Anova reveals that this variable has an effect on p-scores ($F = 92.46, p < 0.2$) Tukey’s HSD indicated

that listeners are more likely to perceive a prominence where there is both a strong and weak syntactic break, but unanimously agree in not perceiving words to be prominent.

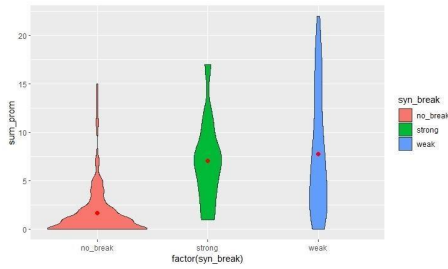


Figure 3: *p*-score as a function of ‘syntactic break’ Scores here (and in subsequent graphs) are shown out of 26, the number of participants.

An examination of whether listeners are affected by the cue of ‘part of speech’, as seen in Figure 4, reveals that listeners are more likely to perceive a prominence when the word is a noun, an adjective and an adverb. These observations were confirmed via a one-way ANOVA ($F = 27.14$, $p < 0.2$) and post-hoc Tukey tests.

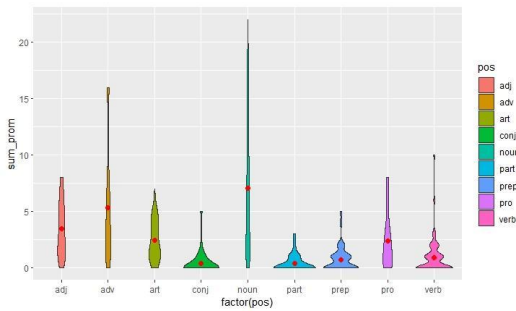


Figure 4: *p*-score as a function of ‘part of speech.’

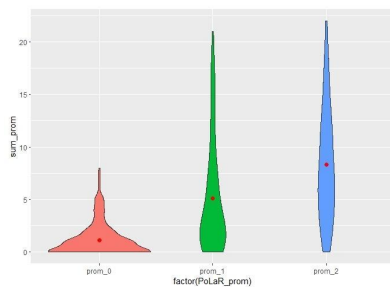


Figure 5: *p*-score as a function of ‘PoLaR prominence labels,’ coded here as *prom_0*, *prom_1* and *prom_2*.

Judgements from listeners about prominence were also correlated with the PoLaR annotations, such that words not marked as prominent in PoLaR were similarly rarely marked as prominent in the RPT task. For words labeled with a prominence (*prom_1*) or strong prominence (*prom_2*), listeners were more likely to mark that word as prominent, with “strong prominence”-marked words showing the highest average prominence scores (see Figure 5.) These differences were reflected via ANOVA ($F = 104.1$, $p < 0.2$) and Tukey HSD tests, showing significant differences across all 3 levels.

3.2. Boundary Experiment Results

The main result to note about the boundary experiment is that our participants had a higher inter-rater agreement for boundary ratings than for prominence ratings. The Fleiss’ kappa score for the boundary experiment amounts to 0.58 ($z = 42.1$), a result consistent with other languages as well (Cole & Shattuck-Hufnagel, 2016; Riesberg et al, 2020). Table 2 compares the inter-rater scores for boundaries for German (Baumann & Winter, 2015), English (Cole et al. 2010), Papua Malayan (Riesberg et al, 2020) and Albanian.

Table 2: Fleiss’ kappa scores for boundaries in German, English, Papua Malay and Albanian

	German	English	PM	Albanian
Fleiss’ Kappa	0.53	0.54	0.40	0.58

Differently from perceptions of prominence, in judging boundaries, listeners’ b-scores correlate only with weak syntactic breaks, but not strong syntactic breaks. These differences were also confirmed via ANOVA ($F = 151.1$, $p < 0.2$) and Tukey HSD tests. In other words, listeners rarely indicate that there is a boundary when there is no syntactic break, but also rarely at strong syntactic breaks, which tend to correspond with the end of the audio file. (These results are shown in Figure 6.)

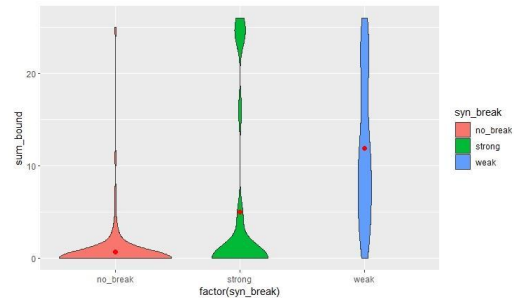


Figure 6: *b*-score as a function of ‘syntactic break.’ Scores here are shown out of 26, the number of participants.

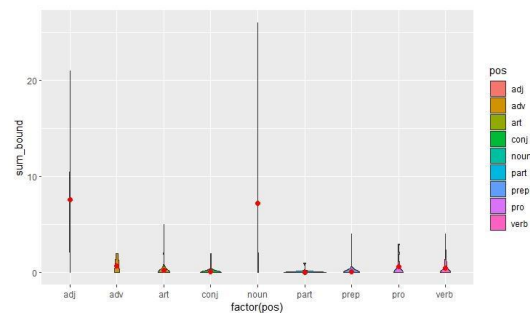


Figure 7: *b*-score as a function of ‘part of speech’

In terms of ‘parts of speech’, listeners were more likely to perceive a boundary, at least for this dataset, after a noun and an adjective, as seen in Figure 7. These observations were also confirmed via ANOVA ($F = 19.6$, $p < 0.2$) and Tukey HSD tests. Figure 8, below, shows significant results comparing RPT boundary scores and PoLaR-labeled boundaries, as indicated by one-way

ANOVA ($F = 131.7, p < 0.2$) and Tukey’s HSD tests. Where the trained labeller marked no phrase-level boundaries, RPT listeners also largely did not mark a boundary. For boundary levels 1 and 2, b-scores increased. Somewhat unexpectedly, the words marked with the largest PoLaR boundary showed very low b-scores. Examination of these data showed that these were largely final words of the file, and listeners may simply not have considered marking them.

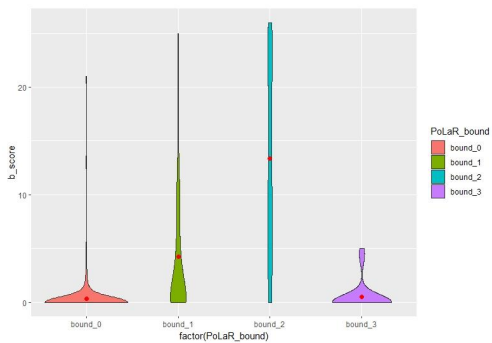


Figure 8: *b-score as a function of ‘PoLaR boundary labels’*

4. Discussion and Conclusion

The results presented in the proceeding section have a number of implications both for investigating the prosodic system of Albanian, and for methodological issues with RPT.

Results from the prominence perception experiment show that, despite great variability among Albanian listeners, they are not in fact stress ‘deaf’. Albanian listeners appear to pay attention to prominences comparably to German and English listeners (e.g. Cole et al, 2010; Baumann & Winter, 2015; Riesberg et al, 2020). In doing so, their prominence judgments correlate with the factors we tested here, i.e. intonational categories of prominence as coded by a trained labeller using PoLaR, syntactic break and part of speech. These results suggest that listeners may be sensitive to these factors in their prominence perception. In the future, we plan to investigate the specific tonal cues to the prominence types coded with PoLaR, such as finding support for pitch accent categories in Albanian discussed in other recent work (Kapia et al, 2020; Kapia et al, forthcoming). Our ongoing work explores other likely acoustic correlates of prominence, including voice quality and duration. (See Kapia et al, 2022, for a study of duration.)

Additionally, Albanian listeners’ perception of prominence seems to be further correlated with non-prosodic factors, which also suggest that these may play a role in perception. Specifically, we saw that prominence perception is correlated with the type of syntactic break, and part of speech. While results about part of speech may be idiosyncratic to these particular audio files and their repetitive format, the results generally point to a direction familiar for many other languages (e.g. Baumann et al, 2016; Roy et al, 2017), in which it has been reported that content words are more prominent than function words, due to their higher semantic weight or structural strength (Buring, 2012) and, commonly, due to their lower word frequency (Cole et al, 2010).

Turning now to boundary perceptions, this study shows that Albanian listeners behave similarly to listeners of other

languages (e.g. Mo & Cole, 2010; Smith, 2011; Jyothi et al, 2014), i.e. they show moderate to substantial agreement in their boundary judgments. In doing so, their boundary perceptions correlate with PoLaR boundary types, syntactic break and part of speech. Indeed, their boundary perceptions correlate with two levels of phrasing between the word level and the intonational phrase level (which we hear call bound_1 and bound_2). We interpret this as support for the previously proposed accentual phrase in Albanian (Kapia et al, 2020; Kapia et al, forthcoming), as well as a potential intermediate phrase level. While we found virtually no agreement from RPT listeners with the trained labeller about the strongest boundaries (bound_3), we believe this to be an artifact of the task; listeners may have interpreted the instructions about grouping as being primarily to divide the utterance into groups, and thus may not have felt compelled to mark the end of the final group. These findings together provide pointers for future investigation phrase levels in Albanian.

Similarly to the prominence perception experiment, listeners’ judgments about boundaries also correlate with weak syntactic boundaries, i.e. syntactic phrases within the audio excerpt that often correlate with relative clauses or prepositions phrases, as well as nouns and adjectives. We again observe a pattern reported for other languages, with boundaries being perceived more often with content words than with function words (Buring 2012; Roy et al, 2017).

Methodologically, this study has shown that while RPT is quite reliable in providing important clues about the prosodic system of a little-investigated language, such as Albanian, its instructions need to be administered carefully with respect to the end of the audio files’ judgements. More specifically, our results indicate that listeners did not think of marking the end of the audio files with regards to boundary perception, despite instructions ‘to mark all the groupings (which could be more than one)’. This was suggested by the fact that strong syntactic breaks, which we would have expected to be correlated with strong boundary perception, had no effect on listeners’ boundary judgements. Notably, such cases coincided with the end of the audio file. Similarly, boundaries coded as the strongest by a trained labeller using PoLaR also coincided with the end of the audio file, and likewise had no effect on listeners’ judgment of boundaries. Possible ways to resolve this issue are to exclude judgments for the last word, as done in previous RPT studies (e.g. Riesberg et al, 2018, 2020), or to include excerpts where the end of the audio file does not coincide with the end of an utterance. Alternately, listeners could be explicitly instructed to consider whether the final words of a file might also be the end of a group.

All in all, while more factors should be investigated in the future, this preliminary analysis suggests that the picture of Albanian prosodic structure and perception is quite complex; a number of different prosodic and non-prosodic factors potentially influence listeners’ judgments of prominences and boundaries. We saw clear correlations of syntactic break, part of speech and the PoLaR-coded prominences and boundaries with naïve listeners’ judgments.

5. Acknowledgements

Many thanks to our participants and student assistants. This research was funded by Alexander von Humboldt Foundation and European Research Council Grant no. 742289 *Human interaction and the evolution of spoken accent* (2017-2022).

6. References

- [1] Ahn, B., Veilleux, N., Brugos, A., & Shattuck-Hufnagel, S. (2019). PoLaR Annotation Guidelines. <https://doi.org/10.17605/OSF.IO/USBX5>
- [2] Baumann, S., Eckart, K., & Riester, A. (2016). Quantifying Prosodic Prominence for Research in Information Structure. Poster presentation. Prosody and Information Structure, Stuttgart. <https://kups.uni-koeln.de/9810/>
- [3] Baumann, S., & Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics*, 70, 20–38. <https://doi.org/10.1016/j.wocn.2018.05.004>
- [4] Boriçi, L. (1987). Intonacioni dhe funksionet e tij në gjuhën shqipe. *Studime Filologjike*, 2.
- [5] Buring, D. (2011). Focus and Intonation. In G. Russell & D. Graff Fara (Eds.), *Routledge Companion to the Philosophy of Language*. Routledge Publishing.
- [6] Cole, J., Mo, Y., & Baek, S. (2010). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes*, 25(7–9), 1141–1177. <https://doi.org/10.1080/01690960903525507>
- [7] Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, 1(2). <https://doi.org/10.1515/labphon.2010.022>
- [8] Cole, J., & Shattuck-Hufnagel, S. (2016). New Methods for Prosodic Transcription: Capturing Variability as a Source of Information. *Laboratory Phonology*, 7(1), 8. <https://doi.org/10.5334/labphon.29>
- [9] Gjinari, J. (1988). *Dialektologjia shqiptare* (revised and expanded). Akademia e Shkencave e Shqipërisë.
- [10] Jyothi, P., Cole, J., Hasegawa-Johnson, M., & Puri, V. (2014). An investigation of prosody in Hindi narrative speech. *Speech Prosody* 2014, 623–627. <https://doi.org/10.21437/SpeechProsody.2014-112>
- [11] Kapia, E., Harrington, J., & Kleber, F. (forthcoming). An Autosegmental-Metrical Analysis of Albanian Prosody. In *Prosody Typology: Vol. III*. Oxford University Press.
- [12] Kapia, E., Kleber, F., & Harrington, J. (2020). An Autosegmental-Metrical Analysis of Rising Contours in Standard Albanian. 10th International Conference on Speech Prosody 2020, 171–175. <https://doi.org/10.21437/SpeechProsody.2020-35>
- [13] Kapia, E., Kleber, F. & Brugos, A. (2022). Discrete and continuous-valued prosodic cues to prominence perception in Albanian. To be presented at *Prominence in Language 2022*, Cologne.
- [14] Kostallari, A. (1984). Gjuha e sotme kombëtare shqipe dhe epoka jonë. *Studime Filologjike*, 21(4), 25–59.
- [15] R Core Team. (2021). R: A language and environment for statistical computing. <https://www.R-project.org>
- [16] Riesberg, S., Kalbertodt, J., Baumann, S., & Himmelmann, N. P. (2020). Using Rapid Prosody Transcription to probe little-known prosodic systems: The case of Papuan Malay. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 11(1), 8. <https://doi.org/10.5334/labphon.192>
- [17] Roy, J., Cole, J., & Mahrt, T. (2017). Individual differences and patterns of convergence in prosody perception. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 8(1), 22. <https://doi.org/10.5334/labphon.108>
- [18] Skopeteas, S. (Ed.). (2006). *Questionnaire on information structure (QUIS): Reference manual*. Universitätsverl.
- [19] Smith, C. (2011). Naïve listeners' perceptions of French prosody compared to the predictions of theoretical models. *Proceedings of the 3rd International Conference on the Discourse-Prosody Interface*, 335–349.