Prosodic and Non-Prosodic Cues to Prominence and Boundaries: Evidence from an RPT study in Albanian

Alejna Brugos[1], Enkeleida Kapia[2]

[1]Simmons College, Boston; [2]Institute for Phonetics & Speech Processing, Munich & ASA, Tiranë

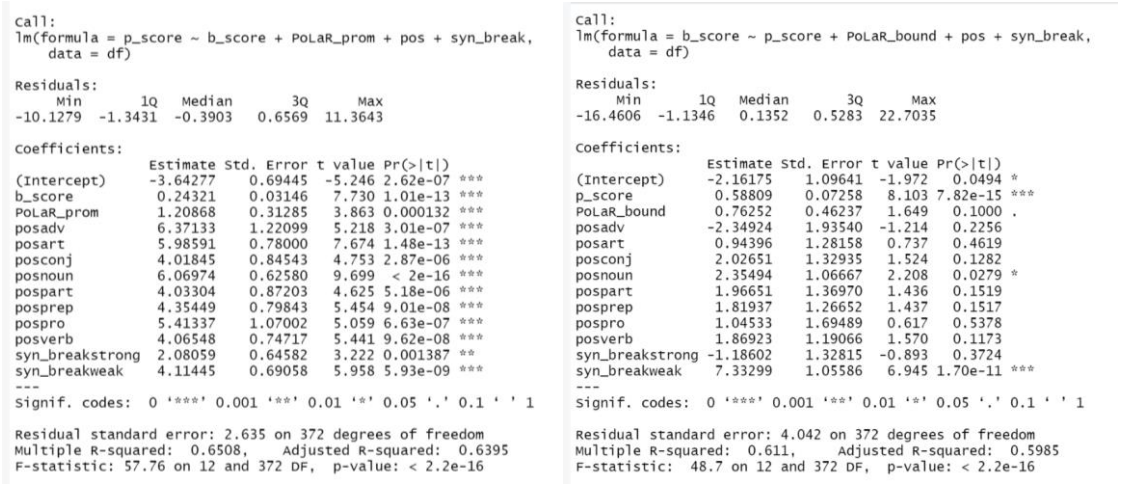
This study reports on a preliminary exploration of prominence and boundaries in Albanian and various prosodic and non-prosodic factors affecting both of them. Using Rapid Prosody Transcription (RPT) [1], we investigated how native Albanian speakers perceive prosodic prominence and boundaries. [2,3] have suggested that prominence in Albanian is marked both by the head and edge of the phrase, i.e. usually a pitch accent associates to the syllable with primary lexical stress and a boundary tone associates to the word's right boundary. Following this, our study expands on these results: 1) by investigating spontaneous speech collected during a story-telling task (vs. read speech), 2) by exploring perception (vs production), 3) by looking at the effect of morpho-syntactic cues on native listeners' perception of prominence or boundaries (vs. just prosodic cues).

Our analysis for this perception study used productions from two male and two female speakers from a large corpus of 45 native speakers. Twenty short (~20-second) sound files were taken from longer interviews in which speakers described 2-picture story sequences (from QUIS [4]). 20 naive listeners participated in the online study via Percy [5]. Following RPT methods, a text transcript of each file was displayed, with words separated by spaces, but no punctuation. Participants listened and clicked on words if they perceived them as highlighted in relation to surrounding words or if they perceived a boundary after them. For each word, two continuous-valued prosody features were calculated: the proportion of transcribers who marked the word as prominent (the p-score, between 0 and 1) and those who marked a boundary (the b-score, between 0 and 1).

We tested the role of a number of prosodic and morpho-syntactic cues that have been found to have an effect on native listeners' perception of prominence and boundaries in other languages [e.g. 6, 7, 8]. More specifically, we looked at the role that intonational categories of prominence and boundaries, present in our PoLaR style intonation analysis [9], had on p- and b-scores, as well as whether p-scores affected b-scores and vice versa. As far as morpho-syntactic cues, we investigated only two for this study, i.e. part of speech and syntactic break (with a future aim to look at more prosodic and morpho-syntactic cues).

Fleiss' K scores were calculated and showed that agreement on b-scores was higher (kappa = 0.58) than agreement on p-scores (kappa = 0.24), a result also found in many RPT studies so far [10]. Furthermore, multiple regression analyses indicated that p-scores could be predicted from b-scores, PoLaR prominence and boundary labels, as well as part of speech and syntactic breaks (as shown in Fig 1 left). b-scores, on the other hand, could only be predicted from p-scores, part of speech and syntactic break at the small ip level (Fig. 1 right). The fact that participants did not perceive a boundary at the big IP level is related to the fact that listeners never thought of marking the end of an utterance, given instructions to detect boundaries in the utterance. Importantly, however, listeners were very consistent about perceiving a boundary whenever they perceived a prominence and vice versa. Lastly, the two morpho-syntactic cues served as strong predictors of both p-scores and b-scores.

In sum, by looking at spontaneous speech using a novel methodology, we have shown here that Albanian listeners perceive a boundary whenever they perceive a prominence and take their cues for this perception from a number of prosodic and non-prosodic events. This study not only illustrates that the RPT method is a helpful tool to probe into the prosodic system of a language, similar to [11], but also suggests that further work needs to be done in order to better understand the interrelation between boundary and prominence perception, as well as the various prosodic and morpho-syntactic factors affecting them.



```
Call:
lm(formula = p_score ~ b_score + PoLaR_prom + pos + syn_break,
    data = df)

Residuals:
     Min       1Q   Median       3Q      Max
-10.1279  -1.3431  -0.3903   0.6569  11.3643

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     -3.64277    0.69445  -5.246 2.62e-07 ***
b_score          0.24321    0.03146   7.730 1.01e-13 ***
PoLaR_prom       1.20868    0.31285   3.863 0.000132 ***
posadv           6.37133    1.22099   5.218 3.01e-07 ***
posart           5.98591    0.78000   7.674 1.48e-13 ***
posconj          4.01845    0.84543   4.753 2.87e-06 ***
posnoun          6.06974    0.62580   9.699  < 2e-16 ***
pospart          4.03304    0.87203   4.625 5.18e-06 ***
posprep          4.35449    0.79843   5.454 9.01e-08 ***
pospro           5.41337    1.07002   5.059 6.63e-07 ***
posverb          4.06548    0.74717   5.441 9.62e-08 ***
syn_breakstrong  2.08059    0.64582   3.222 0.001387 **
syn_breakweak    4.11445    0.69058   5.958 5.93e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.635 on 372 degrees of freedom
Multiple R-squared:  0.6508,    Adjusted R-squared:  0.6395
F-statistic: 57.76 on 12 and 372 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = b_score ~ p_score + PoLaR_bound + pos + syn_break,
    data = df)

Residuals:
     Min       1Q   Median       3Q      Max
-16.4606  -1.1346   0.1352   0.5283  22.7035

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     -2.16175    1.09641  -1.972   0.0494 *
p_score          0.58809    0.07258   8.103 7.82e-15 ***
PoLaR_bound      0.76252    0.46237   1.649   0.1000 .
posadv          -2.34924    1.93540  -1.214   0.2256
posart           0.94396    1.28158   0.737   0.4619
posconj          2.02651    1.32935   1.524   0.1282
posnoun          2.35494    1.06667   2.208   0.0279 *
pospart          1.96651    1.36970   1.436   0.1519
posprep          1.81937    1.26652   1.437   0.1517
pospro           1.04533    1.69489   0.617   0.5378
posverb          1.86923    1.19066   1.570   0.1173
syn_breakstrong -1.18602    1.32815  -0.893   0.3724
syn_breakweak    7.33299    1.05586   6.945 1.70e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.042 on 372 degrees of freedom
Multiple R-squared:  0.611,     Adjusted R-squared:  0.5985
F-statistic:  48.7 on 12 and 372 DF,  p-value: < 2.2e-16
```

Fig 1. Multiple regression model with p_score as response variable (left); multiple regression model with b-score as response variable (right)

Bibliography

[1] Cole, J., & Shattuck-Hufnagel, S. (2016). New methods for prosodic transcription: Capturing variability as a source of information. *Laboratory Phonology,* 7(1): 8, 1–29. DOI: https://doi.org/10.5334/labphon.29

[2] 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.

[3] Kapia, E., Harrington, J., Kleber, F. (forthcoming). An Autosegmental Metrical Analysis of Albanian Prosody. *Prosodic Typology III*, eds. Jun, S. A. Oxford University Press.

[4] Skopeteas, S., Fiedler, I., Hellmuth, S., Schwarz, A., Stoel, R., Fanselow, G., & Krifka, M. (2006). *Questionnaire on information structure (QUIS): reference manual* (Vol. 4). Universitätsverlag Potsdam.

[5] Draxler, C. (2011). Percy – an HTML5 Framework for Media Rich Web Experiments on Mobile Devices. *Interspeech 2011*, 3339–40, Florence, Italy.

[6] Cole, J., Mo, Y., & Baek, S. (2010a). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes*, 25, 1141–1177. DOI: https://doi.org/10.1080/01690960903525507

[7] Rietveld, T. C. M., & Gussenhoven, C. (1985). On the relation between pitch excursion size and pitch prominence. *Journal of Phonetics*, 15, 273–285. DOI: https://doi.org/10.1016/S0095-4470(19)30571-6

[8] Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, 35, 445–472. DOI: https://doi.org/10.1016/j.wocn.2006.12.001

[9] Ahn, B., Veilleux, N., Shattuck-Hufnagel, S. & Brugos, A. (2019). *PoLaR Annotation*. Accessible at http://www.polarlabels.com.

[10] Cole, J., & Shattuck-Hufnagel, S. (2016). New methods for prosodic transcription: Capturing variability as a source of information. *Laboratory Phonology,* 7(1): 8, 1–29. DOI: https://doi.org/10.5334/labphon.29

[11] 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. DOI: http://doi.org/10.5334/labphon.192