Dialect features in heterogeneous and homogeneous Gheg speaking communities

Josiane Riverin-Coutlée[1], *Conceição Cunha*[1], *Enkeleida Kapia*[1,2], *Jonathan Harrington*[1]

[1]Institute of Phonetics and Speech Processing, Ludwig-Maximilians-Universität Munich
[2]Academy of Albanological Sciences
josiane.riverin|cunha|enkeleida.kapia|jmh@phonetik.uni-muenchen.de

Abstract

This apparent and real time study analyses how dialect features in the speech of children and adults are differently affected depending on whether they live in homogeneous or heterogeneous speech communities. The general hypotheses are that speakers in such high contact settings as heterogeneous urban centers are more prone to innovation than speakers in homogeneous tightly-knit communities, and that children accelerate leveling, especially through schooling and socialization. This study is of Gheg Albanian, a dialect spoken in and around the capital Tirana. Two features were investigated: rounding of /a/ and vowel length contrasts. Two groups of adults and children were compared: one from Tirana and one from a nearby village. Additionally, the children were recorded twice over a period of 12 months and were compared longitudinally. The results showed that length contrasts were still present in both communities and age groups. Rounding of /a/ was lost in the city, but undergoing change in the village, with differences measured in apparent time, but also in child speech within the 12-month span. Our study further raises the issue of combining both apparent and real time data within the same design.

Index Terms: Albanian, dialect contact, real and apparent time, acoustics.

1. Introduction

Albanian (*shqip* in Albanian) is a language of the Indo-European family spoken by 6-7 million people [1] living mostly in Albania and Kosovo, but also in North Macedonia, Italy, Greece, Montenegro and Serbia. The two main Albanian dialects, Gheg and Tosk [2], differ on an array of linguistic features, and are used alongside a predominantly Tosk-based standard variety [3]. The focus of this paper is on two dialect features that characterize the variety of Gheg traditionally spoken in and around the capital city of Tirana: contextual rounding of the low vowel /a/ and phonemic vowel length contrasts [2,4,5,6, etc.]. We investigate how these features are realized by children and adults who live in the homogeneous rural community of Bërzhitë versus heterogeneous urban Tirana.

Tirana has been chosen as the locus of our investigation because of its heterogeneous nature: it underwent considerable socio-demographic changes during and after communism. More specifically, a strict spatio-economic planning policy during the communist era turned urban centers such as Tirana into large industrial complexes via interregional allocation of workforce, which meant that workers from around the country were sent there to “serve the country” [7,8, etc.]. In Tirana, this created a high contact situation between the local Gheg speaking community and Tosk speaking migrant workers. Additionally, in post-communist Albania, Tirana received *en masse* internal migrants in search of better economic opportunities, causing the city to triple in size [9] and to experience further linguistic contact. Last but not least, with the city being the country’s main administrative, educational, cultural and media hub, Tirana’s speech community has unquestionably been under a strong influence of the standard variety.

By contrast, the community of Bërzhitë, even though located about 15 km away from Tirana, did not experience anything similar. During communism, as a result of the regime’s spatio-economic policy of retaining the agrarian nature of the countryside [8,10, etc.], villages may have experienced emigration, but no immigration. After communism, a rural exodus took place [11]; again, people moved out, but not into villages. This kept the communities quite homogenous and linguistic contacts to a minimum; often the only outsiders coming into villages like Bërzhitë were the teachers who spoke the standard variety, to which the community was already exposed through public media.

Various impressionistic studies have suggested dialect leveling in different Gheg-speaking cities attributable to the aforementioned high contact situation [7,10]. Our recent study [12] comparing adult Gheg speakers living in Tirana and Bërzhitë has confirmed that rounding of /a/ is almost non-existent in Tirana, while it is being variably produced in the village; phonemic length contrasts, on the other hand, are preserved in both locations. Rounding of /a/ (henceforth, $[a]_{+round}$) refers to the stressed low vowel /a/ being labialized [4] as [ɔ] or [o] when preceded by a nasal consonant, as in *nata* ‘night’ [ˈnɔta]. Loss of this feature means that speakers pronounce [ˈnata] just as in Tosk and the standard (henceforth, Tosk/Standard). Length contrasts are phonemic in Gheg and function primarily as a morpho-phonological marker of indefiniteness, as in: *veza* ‘the egg’ [ˈveza] vs. *një vezë* ‘an egg’ [veːz] [2]. By contrast, both Tosk/Standard only have short vowels.

In the present study, we compare how these two features are produced by adult and child speakers. This cross-generational comparison forms part of an apparent time design and is often used as a proxy for establishing whether a linguistic community is experiencing sound or language change [13,14]. In the dialect contact literature, children are also thought to be instrumental in pushing forward ongoing changes, either because they tend to reanalyze and generalize features they are exposed to, or because they come in contact with peers who do not share the phonetic characteristics of their family [15]. By comparing children and adults, we thus seek to establish whether change in Gheg spoken in the city has further progressed in the direction of Tosk/Standard for the rounding of /a/, and whether it has been initiated for length.

We also added a real time component [14,16] to the design of the study in order to observe finer longitudinal change over time by re-recording the children at yearly intervals (two sets of recordings are completed and analyzed here). We selected first-graders because it has been shown that the onset of schooling and socialization outside the family nucleus may precipitate sound or language change [17,18].

2. Methods

2.1. Participants and Materials

Forty-seven (47) speakers participated in this study, including 26 children and 21 adults. The children, 14 girls and 12 boys, were recorded twice over a period of 12 months, first when they were in first grade, aged 6-7 years old, and second when they were in second grade. Fifteen (15) children were from Tirana and 11 from Bërzhitë. The 21 adults, all women aged 29-72 years old ($\bar{x}$ = 43), were approached via the children's primary schools; they were mothers, grandmothers or acquaintances recruited through snowball sampling. Eleven (11) adults were from Tirana and 10 from Bërzhitë. All participants were native Gheg speakers born and raised in the same Gheg-speaking area.

The recordings took place in quiet rooms of primary schools in Tirana and Bërzhitë with a Beyerdynamic TG H54c head-mounted microphone. The speech signal was digitally recorded (44,100 Hz, 16 bits) with a Tascam US-2x2 and the Speech Recorder software [19].

As the children could not yet read in first grade, the participants took part in a picture naming task featuring 54 different items, 29 of which are analyzed here. Because the lexical items chosen had to be depictable in images and meaningful to children, it was not possible to control for such factors as stress pattern, number of syllables, and phonetic context. Each lexical item was produced four times per speaker per recording session (7924 tokens after removal of unanalyzable material).

The speech signal was forced-aligned using WebMAUS [20,21] and structured into a speech database using EMU-SDMS [22]. The statistical analyses were carried out in R, using the *lme4* and *emmeans* packages [23,24,25].

2.2. Contextual rounding

Eleven words were analyzed: 3 with [a]+round, 4 with [a]−round and 4 with /o/ vowels. Formant frequencies were estimated with linear prediction coding (LPC) using the Burg algorithm. Five formants were calculated in the 0-7000 Hertz range for children, and 0-5500 for female adults. For each vowel, F1 and F2 calculated at 11 equally spaced time points between vowel onset and offset were smoothed with a 5-point median filter. In order to parameterize the dynamic shape of each formant trajectory, F1 and F2 were decomposed into sets of half-cycle cosine waves using the discrete cosine transformation (DCT) resulting in three coefficients, k_0, k_1 and k_2, for each formant (thus 6 coefficients, 3 for F1 and 3 for F2 per vowel) that are proportional respectively to the formant's mean, linear slope, and curvature. The normalised orthogonal projection *op* of a vowel $\vec{x}_S$ in a six-dimensional space formed by the DCT coefficients was calculated in order to determine the relative distance of any [a]+round vowel to the same speaker's [a]−round and /o/ from (1):

$$op(\vec{x}_s) = 1 - 2\frac{(\vec{x}_s - \vec{c}_a)\odot(\vec{c}_a - \vec{c}_o)}{(\vec{c}_a - \vec{c}_o)\odot(\vec{c}_a - \vec{c}_o)} \quad (1)$$

in which $\vec{x}_S$ is the position (vector of 6 values) of any individual [a]+round vowel in a six-dimensional space formed by the DCT coefficients, $\vec{c}_a$ and $\vec{c}_o$ are the centroids (means) of all [a]−round and /o/ vowels produced by the same speaker in the same DCT space, and $\odot$ is the scalar (inner) product of two vectors [26]. Formula (1) expresses in a single value the relative proximity of a given vowel's combined F1 and F2 trajectories (encoded as DCT coefficients) to the mean F1 and F2 trajectory shapes of [a]−round and /o/: the closer [a]+round is in this acoustic space to [a]−round or /o/, the closer the values of this parameter are to +1 and −1 respectively. The output of (1) was the dependent variable in two separate linear mixed-effects regression models, one comparing adults and children (apparent time, formula 2), and one assessing intra-individual change in children (real time, formula 3):

lmer(response ~ Year + AgeGroup*Origin + (1|Speaker) + (AgeGroup+Origin|Word)) (2)

lmer(response ~ Year*Origin + (Year|Speaker) + (Origin+Year|Word)) (3)

where *AgeGroup* is a two-level fixed factor (Adults/Children), *Origin* is a two-level fixed factor (City/Village), and *Speaker* and *Word* are random factors. Also note that in (2), the variable *Year* models the variance explained by the longitudinal component, but is not of interest per se. It has two levels, Year1 and Year2, with adults coded as Year1. Since it has an asymmetrical structure (no adult is represented in Year2, but all children are), it is not set as a slope for the random effect *Speaker*. Formula (3) excludes adults such that the two levels of Year correspond to Year1 and Year2 of data from children.

For Figure 1 only, F1 and F2 at the midpoint of the vowels were speaker-normalized as z-scores [27]. In order to do so, we used 4 words with /i/, 3 with /y/, 5 with /u/ and 3 with /e/ in addition to the 11 words with /o/, [a]+round and [a]−round already mentioned.

2.3. Phonemic length

Twenty-five words were analyzed,19 of which contained long vowels and 6 short vowels. Duration was measured after hand-correction of the boundaries marking the onset and offset of the vowels. The statistical models in (4, 5) included the log-transformed duration as the response variable, *Length* as a two-level fixed factor (Long/Short) and the other fixed and random factors as in (2, 3):

lmer(log(response) ~ Year + AgeGroup*Length*Origin + (Length|Speaker) + (AgeGroup+Origin|Word)) (4)

lmer(log(response) ~ Year*Length*Origin + (Length+Year|Speaker) + (Origin+Year|Word)) (5)

3. Results

3.1. Contextual rounding

Figure 1 shows how tokens of [a]+round are distributed in F1/F2 planes compared to the speakers' baseline [a]−round and /o/, represented by gray ellipses. Figure 1 also includes a series of density plots illustrating how the values of the orthogonal projection (*op*) of [a]+round are distributed relative to the anchors [a]−round corresponding to 1, and /o/ to −1. We observe that the [a]+round tokens produced by the city adults overlap mostly with their [a]−round ellipse and have *op* values around

+1. The village speakers' tokens overlap with both $[a]_{-\text{round}}$ and /o/ in the F1/F2 plane and have a quasi-bimodal distribution of *op* values with peaks centered around +1 and −1, suggesting the presence of both rounded and unrounded variants, consistently with an ongoing change [14].

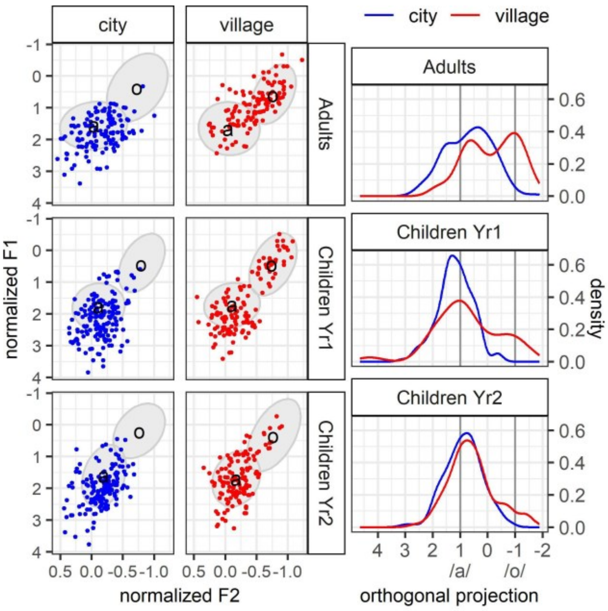

Figure 1: *Tokens of $[a]_{+\text{round}}$ projected on baseline $[a]_{-\text{round}}$ (coded /a/) and /o/ ellipses in normalized F1/F2 space (left & middle); distribution of values of orthogonal projection, where $[a]_{-\text{round}}$ equals 1 and /o/ equals −1 (right)*

The city children, like the city adults, mainly produced tokens overlapping with $[a]_{-\text{round}}$, but the distribution of their *op* values is more condensed. They also appear quite stable from Year1 to Year2. The village children, on the other hand, produced in Year1 both rounded and unrounded variants, but fewer rounded ones than the adults, as reflected in the F1/F2 plane and distribution, where the peak around −1 is smaller than the +1 peak. In Year2, they produced even fewer rounded variants and were much more similar to city children than to village adults.

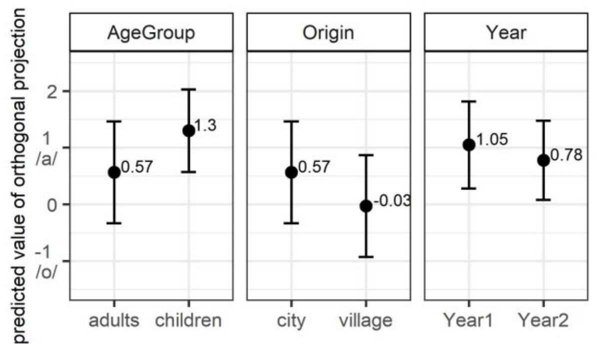

Figure 2: *Predicted value of orthogonal projection for the levels of the significant effects in models (1) and (2), where baseline $[a]_{-\text{round}}$ (coded /a/) equals 1 and /o/ equals −1*

The results of the statistical analyses showed a significant influence of both AgeGroup ($t[9.7] = 5.02$, $p < 0.001$) and of Origin ($t[44.5] = 5.17$, $p < 0.001$) on *op*, but no interaction between these factors. As can be seen in Figure 2, the predicted value is significantly higher for children, meaning that their tokens are closer to $[a]_{-\text{round}}$ than they are in adults. They even seem to be more peripheral than $[a]_{-\text{round}}$, as can be inferred from a mean predicted value over +1. Furthermore, the predicted *op* value is lower for village speakers, i.e., at around 0. This likely reflects the averaging of the +1 and −1 peaks observed in Figure 1. The results of the second model applied to the data from children only reveal a significant effect of Year ($t[17.6] = 2.27$, $p = 0.035$), but not of Origin, and no interaction between these factors. Figure 2 shows that the predicted value at Year2 is lower than at Year1, suggesting that the children's productions have become less peripheral in the vowel space over time.

3.2. Phonemic length

Figure 3 displays duration measured in short and long vowels. It is clear that all groups of speakers produced the length contrast. In general, villagers produced longer vowels than city dwellers, especially in the Long words. Children's vowels were also longer than for adults, possibly due to a slower speech rate [28].

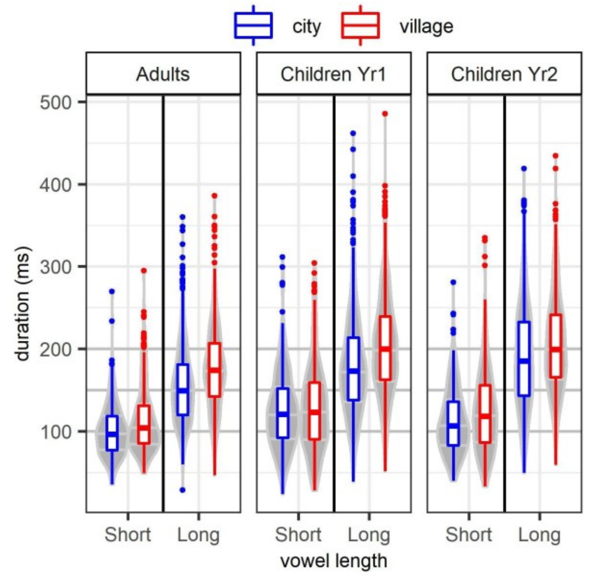

Figure 3: *Duration of vowels in short and long words*

The results of the statistical analyses showed a significant influence on vowel duration of AgeGroup ($t[58.9] = 2.64$, $p = 0.009$), Origin ($t[55.2] = 2.37$, $p = 0.020$) and Length ($t[27.5] = 6.72$, $p < 0.001$), with no interaction between these factors. As shown in Figure 4, the first model predicts greater durations for child and village speakers, irrespective of expected length. A greater duration is also predicted for the long vowels.

The model fitted to the child data shows a significant three-way interaction between Year, Origin and Length ($t[4403] = 4.52$, $p < 0.001$). As evidenced by the results of the post-hoc comparisons, presented in Figure 5, village children remained quite stable over time and the difference between their long and short vowels is predicted to be greater than for the city children. The latter changed over time, with a slightly more marked difference between short and long vowels in Year2. The duration of their short vowels, in particular, has become closer to adult values (see Figure 3). Beyond the three-way interaction, the only significant effect is that of Length ($t[31] = 4.34$, $p < 0.001$).

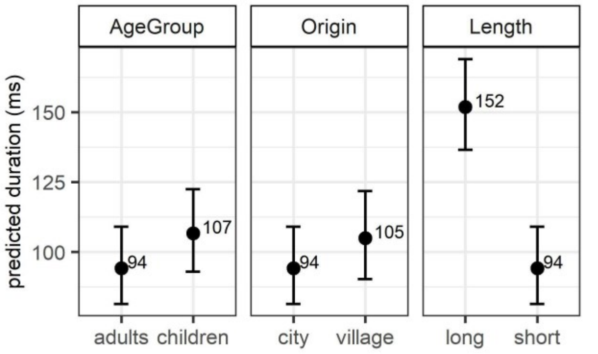

Figure 4: *Predicted duration for the levels of the significant effects in model (3) (exponentiated)*

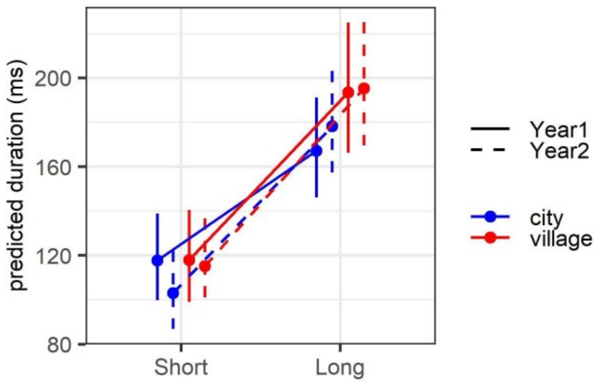

Figure 5: *Predicted duration for the levels of the triple interaction in model (4) (exponentiated)*

4. Discussion and Conclusions

In this study, we compared the attrition of two dialect features of Gheg for Gheg speakers living in a heterogeneous, high contact city (Tirana) as opposed to a more homogeneous, tightly-knit rural community (Bërzhitë). In our earlier study on adults [12], Gheg speakers preserved length contrasts in both these locations, while phonetic rounding of /a/ appeared to be lost in Tirana and also likely undergoing change in Bërzhitë. In this paper, we sought to verify whether leveling was more advanced for children living in these two communities.

For /a/-rounding, we found that whereas village speakers produced variants of $[a]_{+round}$ that were acoustically similar to either $[a]_{-round}$ or /o/ (bimodal distribution), the city speakers' $[a]_{+round}$ was much closer to $[a]_{-round}$. This confirms our previous work, but does not indicate that the children behave differently from the adults (no significant interaction between AgeGroup and Origin). As a whole, child productions were significantly closer to $[a]_{-round}$ than adult productions, but they also got further away from $[a]_{-round}$ over time, in both the city and the village. This could indicate that the children's vocalic category is becoming less peripheral and more adult-like. Figure 2 also showed village children had produced less rounded tokens in Year2 than Year1, but this observation was not statistically supported by an interaction in model (2).

Contrastive length definitely seems to be preserved among our participants. Once again, we found no evidence that children behave differently from the adults, except that their vowels generally had a greater duration, which could come as result of a slower speech rate [28]. Moreover, village speakers produced vowels with a greater duration than city speakers; our qualitative observations suggest this is mostly due to long vowels having an especially large duration in the village, but this was not confirmed statistically. While the village children remained very stable from Year1 to Year2, their city peers slightly increased the contrast between short and long vowels over time, away from Tosk/Standard, similarly to what has been previously reported in [17], where 4–5-year-old children converged toward nonstandard, rather than standard variants.

In sum, in the city, we observe that both adults and children have lost rounded /a/ and preserved length contrasts. In the village, there are signs of change: the adults produced both rounded and unrounded variants of $[a]_{+round}$, while the children produced less rounded variants than the adults in Year1, and even less just a year later. This tends to indicate that children are pushing the ongoing change in the village. Length, on the other hand, may be resisting change even in a heterogeneous and high contact setting like Tirana because it is a morpho-phonological marker, not just a phonetic variant like /a/-rounding, and such linguistically complex features are expected to change more slowly [15,29].

Notably, however, some of the changes we observe from Year1 to Year2 could partly be explained by growth [30,31,32]. For example, the finding that $[a]_{+round}$ has become less peripheral could be due to less hyperarticulated speech as children grow, which in turn may be the reason why we do not observe significant ongoing loss of the feature among village children. Additionally, the increased length contrast in city children in Year2 is also due in part to the duration of their short vowels becoming closer to that of the adults' short vowels, although it is not entirely clear why they should be ahead of their village peers in acquiring adult-like durations. As our study unfolds and additional recordings are made, one of the goals will be to tease apart effects of growth and dialect change.

Limitations of our study include a relatively small number of participants. Even though such low counts are typical of longitudinal work having to make do with sample attrition, this results in less statistical power, which in turn, may mask certain effects. Also, because we used a picture naming task, there are phonetic factors we could not control for that may have increased variability. Perhaps the greatest limitation of this work is that we had to break down the statistical analysis of each feature into two distinct models, one dealing with apparent time and the other with real time. Future work should address the issue of integrating both in a single model, given the added depth such a design brings to our understanding of language variation and change.

5. Acknowledgements

Many thanks to our participants and student assistants. We also thank Michele Gubian (IPS-LMU) for his statistical advice. This research is funded by European Research Council Grant no. 742289 *Human interaction and the evolution of spoken accent* (2017-2022).

6. References

[1] A. Rusakov, "Albanian", in M. Kapović (ed.), *The Indo-European Languages*, 2nd ed. London: Routledge, pp. 552–608, 2017.

[2] J. Gjinari, B. Beci, G. Shkurtaj, Xh. Gosturani, and A. Dodi, *Atlasi dialektologjik i gjuhës shqipe*. Napoli: Università degli Studi di Napoli, 2007.

[3] A. Kostallari, “Gjuha e sotme kombëtare shqipe dhe epoka jonë”, *Studime Filologjike*, vol. 21, no. 4, pp. 25–59, 1984.
[4] M. Çeliku, “Vëzhgime mbi të folmen e krahinës së Peqinit”, *Studime Filologjike*, vol. 2, no. 4, pp. 91–131, 1965.
[5] B. Beci, “Rreth tipareve karakteristike të dy dialekteve të shqipes”, Studime Filologjike, vol. 15, no. 4, pp. 53–87, 1978.
[6] K. Topalli, *Fonetika historike e gjuhës shqipe*. Tiranë: Shtëpia Botuese Dituria, 2007.
[7] G. Shkurtaj, “Shtresëzime dhe risi gjuhësore në qytetin e Lezhës”, *Studime Filologjike*, vol. 21, no. 2, pp. 75–93, 1984.
[8] O. Sjoberg, “Underurbanisation and the Zero Urban Growth Hypothesis: Diverted migration in Albania”, *Geografiska Annaler. Series B, Human Geography*, vol. 74, no. 1, pp. 3–19, 1992.
[9] A. Gjonça, P. Elezi, and L. Sado, *Pabarazitë në Tiranën e madhe*, INSTAT,.www.instat.gov.al. 2015.
[10] G. Shkurtaj, “Rreth ndryshimeve në gjuhën e fshatit të sotëm të Veriut (sipas vëzhgimeve në zonën e Koplikut, rrethi i Shkodrës)”, *Studime Filologjike*, vol. 6, no. 4, pp. 161–169, 1969.
[11] M. Lerch, “Internal and international migration across the urban hierarchy in Albania”, *Population Research and Policy Review*, vol. 35, no. 6, pp. 851–876, 2016.
[12] J. Riverin-Coutlée, E. Kapia, C. Cunha, and J. Harrington, “Vowels in urban and rural Albanian: The case of the Southern Gheg dialect”, in preparation.
[13] W. Labov, *Principles of Linguistic Change: Internal Factors*, Vol. 1. Oxford: Blackwell, 1994.
[14] G. Bailey, “Real and apparent time”, in J. K. Chambers, P. Trudgill, and N. Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, 1st ed. Oxford: Blackwell, pp. 312–332, 2002.
[15] P. Kerswill, “Children, adolescents, and language change”, *Language Variation and Change*, vol. 8, no. 2, pp. 177–202, 1996.
[16] G. Sankoff, and H. Blondeau, “Language change across the lifespan: /r/ in Montreal French”, *Language*, vol. 83, no. 3, pp. 560–588, 2007.
[17] A. Nardy, J.P. Chevrot, and S. Barbu, “Sociolinguistic convergence and social interactions within a group of preschoolers: A longitudinal study”, *Language Variation and Change*, vol. 26, no. 3, pp. 273–301, 2014.
[18] S. A. Tagliamonte, and S. Molfenter, “How’d you get that accent?: Acquiring a second dialect of the same language”, *Language in Society*, vol. 36, no. 5, pp. 649–675, 2007.
[19] C. Draxler, and K. Jänsch, “SpeechRecorder – A universal platform independent multi-channel audio recording software”, in *Proceedings of the 4th International Conference on Language Resources and Evaluation*, Lisbon, Portugal, pp. 559–562, 2004.
[20] T. Kisler, U. D. Reichel, and F. Schiel, “Multilingual processing of speech via web services”, *Computer Speech & Language*, vol. 45, pp. 326–347, 2017.
[21] F. Schiel, “Automatic phonetic transcription of non-prompted speech”, in *Proceedings of ICPhS 14*, San Francisco, USA, pp. 607–610, 1999.
[22] R. Winkelmann, J. Harrington, and K. Jänsch, “EMU-SDMS: Advanced speech database management and analysis in R”, *Computer Speech & Language*, vol. 45, pp. 392–410, 2017.
[23] R Core Team. *R: A language and environment for statistical computing*. https://www.R-project.org. 2021.
[24] D. Bates, M. Maechler, B. Bolker, and S. Walker, “Fitting linear mixed-effects models using lme4”, *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.
[25] R. Lenth, J. Love, and M. Herve, *emmeans: Estimated marginal means, aka least-squares means*, R package, 2018.
[26] M. Stevens, J. Harrington, and F. Schiel, “Associating the origin and spread of sound change using agent-based modeling applied to /s/-retraction in English”, *Glossa*, vol. 4, no. 1, article 8, 2019.
[27] B. M. Lobanov, “Classification of Russian vowels spoken by different speakers”, *Journal of the Acoustical Society of America*, vol. 49, no. 2B, pp. 606–608, 1971.
[28] I. P. Martins, R. Vieira, C. Loureiro, and M. E. Santos, “Speech rate and fluency in children and adolescents”, *Child Neuropsychology*, vol. 13, no. 4, pp. 319–332, 2007.
[29] J. K. Chambers, “Dialect acquisition”, *Language*, vol. 68, no. 4, pp. 673–705, 1992.
[30] S. Lee, A. Potamianos, and S. Narayanan, “Acoustics of children’s speech: Developmental changes of temporal and spectral parameters”, *Journal of the Acoustical Society of America*, vol. 105, no. 3, pp. 1455–1468, 1999.
[31] H. Chung, E. J. Kong, J. Edwards, G. Weismer, M. Fourakis, and Y. Hwang, “Cross-linguistic studies of children’s and adults’ vowel spaces”, *Journal of the Acoustical Society of America*, vol. 131, no. 1, pp. 442–454, 2012.
[32] J. Yang, and R. A. Fox, “Acoustic development of vowel production in American English children”, in *Proceedings of Interspeech 2013*, Lyon, France, 1263–1267, 2013.