

Dialect change and language attitudes in Albania

Josiane Riverin-Coutlée¹, Enkeleida Kapia^{1,2} and Michele Gubian¹

¹Institute for Phonetics and Speech Processing, Ludwig-Maximilians-Universität München, Munich, Germany and ²Albanian Academy of Sciences, Tirana, Albania

Corresponding author: Josiane Riverin-Coutlée. Email: josiane.riverin@phonetik.uni-muenchen.de

Abstract

This study is concerned with attitudes of Albanian listeners toward the two main dialects spoken in Albania: Gheg and Tosk. The study seeks to establish a connection between attitudes and speech features which have been shown to be changing in Gheg, and other features found to be stable. Ratings of four speech features on visual analog scales (VASs) pertaining to dialect identification, status, and solidarity were collected from 125 Albanian listeners and modeled with Bayesian regressions. The results revealed lower status for variants of features found to be changing in Gheg, contrary to stable variants, suggesting a connection between attitudes and dialect change, and highlighting the relevance of both language-external and internal factors in understanding change. All stimuli were also rated as more friendly than unfriendly, which could be related to sociocultural specificities of Albania. The study finally identifies methodological challenges to do with modeling responses from VASs.

Keywords: language attitudes; dialect identification; Albanian; dialect change; visual analog scales

Introduction

This study is concerned with language attitudes of Albanians toward specific phonological and phonetic features of the two main dialects of Albanian: Gheg and Tosk. We seek to establish whether there is a connection between attitudes and the recently documented tendency of certain features of Gheg to be changing while others were found to be stable.

Language attitudes

Dragojevic, Fasoli, Cramer, and Rakić (2021:61) defined attitudes as “evaluative reaction[s] to an object,” and language attitudes as “evaluative reactions to language.” Two orthogonal dimensions of language attitudes have been identified: *status* and *solidarity* (also called *competence* and *warmth*; Fiske, Cuddy, Glick, & Xu, 2002). The basis of perceived status is mainly socioeconomic, whereas the sense of solidarity is driven by loyalty and competition (Ryan, 1983). There is a recurring tendency for

lower-status languages or varieties to rank high with respect to solidarity (Labov, 2006:402; Lambert, Hodgson, Gardner, & Fillenbaum, 1960; Trudgill, 1972), but it is not necessarily the case that high status goes together with low solidarity (e.g., Bishop, Coupland, & Garrett, 2005; Cavallaro & Chin, 2009; Cole, 2021). Members of a speech community usually show high consistency in their judgments toward themselves (in-group) and speakers of other speech communities (out-group), which is taken as evidence that these reflect stereotypes (Dragojevic et al., 2021; Fiske et al., 2002; Ryan, 1983).

The current study uses the “verbal guise” technique (Cooper, 1975; Garrett, 2010:53–87), wherein listeners had to rate stimuli produced by various speakers on visual analog scales (VASs), the poles of which showed pairs of contextual antonyms. A considerable number of antonyms and traits have been tested over the years (e.g., Zahn & Hopper, 1985), as these have to be relevant to the community investigated (Garrett, Coupland, & Williams, 2003; Loureiro-Rodríguez & Acar, 2022), but they can still be grouped into the two dimensions of *status* and *solidarity*. A drawback of indirect approaches like the verbal guise technique is that listeners may complete the task without responding to the categories the researchers are investigating (Dragojevic & Goatley-Soan, 2022; Garrett, 2010:57–58; Preston, 1989:4). For example, it may not be so clear that attitudes are directed at dialects, knowing that in dialect categorization tasks, participants tend to group stimuli into fewer categories than established through research (e.g., Avanzi & Boula de Mareüil, 2019; Clopper & Pisoni, 2004b) and show greater accuracy when using broader than narrower labels, for instance, Scottish English compared to Glasgow English (Braber, Smith, Wright, Hardy, & Robson, 2023). In addition, perception has been found to be biased by beliefs held by participants, including laboratory-induced beliefs about the speakers’ origin (e.g., Hay & Drager, 2010; Niedzielski, 1999). It is thus advisable to complement indirect measures of attitudes with explicit questions about the particular dimension investigated, here with a dialect identification question (Dragojevic & Goatley-Soan, 2022; Preston, 1989:4). The VAS response format constitutes another challenge in itself. While allowing participants to express nuances in their judgments (Llamas & Watt, 2014), VASs are also subject to certain biases (Matejka, Glueck, Grossman, & Fitzmaurice, 2016) and high variability in rating strategies between listeners (e.g., Austen & Campbell-Kibler, 2022; de Hoop, Levshina, & Segers, 2023), leading to modest performance of even the most sophisticated statistical modeling techniques currently available (Liu & Eugenio, 2018).

A branch of attitude studies focuses on evaluative reactions to specific linguistic features (Edwards, 1999). While methodologically challenging (Austen & Campbell-Kibler, 2022; Watson & Clark, 2015) in typical guise experiments that use one- or two-sentence-long stimuli containing multiple linguistic features, identifying evaluative reactions to specific features via alternative methods has shown that these were not equally noticed and evaluated by listeners. For example, interviews designed to elicit metalinguistic comments have highlighted that lay speakers may name, describe, and judge specific features, such as T-glottaling in British English (Alderton, 2020) or vowel nasality in Gheg Albanian (Morgan, 2015; see next section); whereas other features described by linguists are not mentioned in such interviews, for instance, high vowel laxing in Quebec French (Lappin, 1982). Societal treatment studies (Garett, 2010), which exploit material spontaneously produced by the speech community like

judgments and representations conveyed in social or traditional media (Cheshire & Moser, 1994), also illustrate this selective awareness. The implicit association test (see Campbell-Kibler, 2012; Greenwald, McGhee, & Schwartz, 1998), where participants associate a given stimulus with certain labels under the assumption that associations are facilitated when labels are implicitly coherent, may grant access to attitudes that participants would perhaps lack resources to discuss during an interview, or would not *wish* to discuss (see Kristiansen, 2009 and Pharao, *in press* for more about implicit versus explicit attitudes).

Labov (1972:314) characterized this selective awareness under the concepts of *stereotype*, *marker*, and *indicator*. Stereotypes are features noticed by the speech community, explicitly commented upon, triggering a measurable evaluative reaction.¹ Markers also trigger some form of reaction, for example during an implicit association task, but not necessarily explicit and polarized comments, as markers “may lie below the level of conscious awareness” (Labov, 1972:314). Indicators are linguistic features of which the speech community is unaware. Several studies have shown that negatively evaluated features are susceptible to processes such as leveling and standardization (e.g., Auer, 2018; Hiramoto, 2010; Kerswill & Williams, 2002; Kristiansen, 2009; Labov, 1963; Milroy, 2001; Pharao, *in press*; Vaicekauskienė, 2019), though this is not a universal tendency, with some low status languages, varieties, or features also exhibiting “stubborn persistence” (Ryan, 1979:147).

Attitudes toward Albanian: What we know so far

Albanian is a lesser-studied language of the Indo-European family spoken by approximately 7 million people worldwide (Rusakov, 2017). Albanian spoken in Albania, which is the focus of our study, comprises two major dialects: Gheg and Tosk. Gheg is spoken in central and northern Albania, including the capital city Tirana, whereas Tosk is spoken in southern Albania (Gjinari, Beci, Shkurtaç, Gosturani, & Dodi, 2007). The two dialects differ (sometimes sharply) on many lexical, morphosyntactic, phonological, and phonetic features, but usually remain mutually intelligible.

Up until 1972, Albanian could have been described as pluricentric: both Gheg and Tosk were employed in writing, depending on the author’s preference (Byron, 1976:41–76). This changed when the National Congress of Orthography proclaimed a standard variety, in 1972, following the aspiration of the Albanian state to unify the country in the common use of one language (Kostallari, 1970). Most scholars consider that standard Albanian is largely based on Tosk, with only a few peripheral Gheg features adopted (Byron, 1976:59–76; Moosmüller & Granser, 2006). The state took several measures, some coercive, to enforce the use of standard Albanian (see Beci, 2000:58–69; Pipa, 1989). These included a ban on publications in any other variety than the new Tosk-based standard, destruction of documents previously written in Gheg, and imprisonment of those refusing to write in the new standard and their families being sent to camps. Broadcasters and other media figures, government officials, and civil servants were obliged to speak the standard. All educational material was offered only in the standard, teachers were pressured to use it, and children were punished when using dialect features, which particularly impacted Gheg-speaking children. Despite the regime of the People’s Socialist Republic of Albania falling in the

early 1990s, the Tosk-based standard remained, and it is still today the only system officially recognized by the state. The measures taken to enforce its use have certainly had an impact on Albanians' representations of the two dialects and their speakers, and the idea of Tosk being standard and Gheg nonstandard is likely well entrenched in Albanian society.

Scholarly research on attitudes toward Albanian is however scarce. A series of experiments conducted by Dickerson (2021) with heritage and expatriate speakers of Albanian living in the United States included a verbal guise experiment in which sentence-long stimuli comprising 13 phonetic, phonological, morphosyntactic, and lexical features differing between Gheg and Tosk were presented to the participants. Heritage and expatriate participants alike, whichever their dialect background, judged Gheg to sound stronger, less proper, and more rural than Tosk. However, both Gheg and Tosk were judged to be similarly friendly.

Morgan (2015) conducted interviews to elicit attitudes from 19 Tosk speakers having moved from southern Albania to Tirana. When it came to the Gheg-Tosk division, the participants described Gheg with adjectives such as *isolated*, *undeveloped*, *rural*, *backwards*, and *uncultured/thick*. They explicitly mentioned two phonological features of Gheg driving their judgments: vowel nasality and monophthongization.² In contrast, Tosk was described as *more developed*, *standard*, *soft*, and *calm*.

Dialect stability and change in Gheg

Recent work has investigated dialect stability and change in Gheg. Riverin-Coutlée, Kapia, Cunha, and Harrington (2022) analyzed vowels produced by Gheg-speaking adults and first grade children living in urban Tirana and rural Bërzhitë to find out whether Gheg was changing under the influence of standard Albanian and, in the case of Tirana, contact with Tosk as well. Three vowel features differing between Gheg and Tosk were analyzed: rounding of /a/, contrastive vowel length, and monophthongization. They found rounding of /a/ to be the most advanced change, with evidence that speakers from both urban and rural areas were adopting the standard variant. In contrast, contrastive vowel length did not show any sign of change in either location. Monophthongization was at an intermediate stage, having started to change in the urban setting, but not in the rural one. These results were tentatively explained based on the features' relative linguistic complexity: rounding of /a/ had changed first and faster than the other features because of its less complex and allophonic nature. However, Riverin-Coutlée et al. (2022) could not quite explain why monophthongization was at an intermediate stage of change, while contrastive vowel length was completely stable, when their relative complexity suggested that it should have been the other way around. A question that arose from this study was whether the sociolinguistic value of these features could at least partly explain the results that linguistic complexity, a language-internal factor, could not (Riverin-Coutlée et al., 2022:496).

Questions and hypotheses

In this study, we set out to explore the potential role of a language-external factor, namely, attitudes. Our aim is to document the attitudes of Albanians toward

the Gheg and Tosk variants of the three features investigated in Riverin-Coutlée et al. (2022): rounding of /a/, contrastive vowel length, and monophthongization; as well as the vowel nasality feature discussed in Morgan (2015). Our main question is whether these features are evaluated differently; that is, whether the low status reputedly attributed to the Gheg dialect, and the high status attributed to the Tosk dialect (Dickerson, 2021), apply uniformly to their variants of these four features. The answer will shed light on a possible connection between perceived status and dialect stability and change. This main question, which addresses the status dimension, naturally raises that of the solidarity dimension. Moreover, as suggested in the literature, we also seek to validate that the participants' attitudes are effectively directed at dialects.

We hypothesize that Albanians' responses will differ across features, with rounding of /a/, monophthongization, and vowel nasality showing different tendencies than contrastive vowel length. We predict reliable dialect identification for the first three features, and the Gheg variants of these features to be perceived as having lower status than the Tosk variants. Because low status tends to be associated with high solidarity, but also because Dickerson (2021:168) found no difference between solidarity expressed toward Gheg and Tosk, we predict both the Gheg and Tosk variants of these three features to trigger high solidarity. On the other hand, for contrastive vowel length, we predict lower accuracy in dialect identification. We also predict listeners to judge both Tosk and Gheg variants of this feature as having similarly high status and high solidarity. Finally, we expect Albanians of both dialect backgrounds to be consistent in their attitudes toward all features; that is, we predict no difference in response patterns according to dialect background of the listener (Gheg versus Tosk).

Methods

Features

Four features, or variables, were investigated in this study (for more details, see Beci, 1995; Çeliku, 1968, 1971; de Vaan, 2018; Gjinari, 1968; Gjinari et al., 2007). First, rounding of /a/ is a phonetic process where, in Gheg, the low vowel /a/ is rounded into [ɔ] in a stressed syllable whose onset is a nasal consonant (e.g., *mal* 'mountain' [mɔl]). In Tosk and standard Albanian, *mal* is instead realized with an unrounded low vowel: [mal].

Second, contrastive vowel length is a morphophonological feature, where Gheg contrasts long and short vowels while most Tosk-speaking areas and standard Albanian have only short vowels. An example of a word with a long vowel in Gheg is *mi* 'a mouse' /mi:/, which is realized with a short vowel, /mi/, in Tosk and standard Albanian. All Gheg vowels have short and long counterparts, and while length contrasts are restricted to stressed syllables, these can be of various types (e.g., open or closed syllables), and in different positions within words.

Third, monophthongization refers to the production by Gheg speakers of monophthongs in words that have diphthongs (or vowel sequences) in Tosk and standard Albanian. For example, *duar* 'hands' is realized /duɹ/ in Gheg, but /duaɹ/ in Tosk and standard Albanian.

Fourth, vowel nasality is a phonological feature of Gheg that is absent from Tosk and standard Albanian, which only have oral vowels. For example, *hënë* ‘moon’ corresponds to /hãn/ in Gheg, but to /hənə/ (pronounced [hənə] or [hən]) in Tosk and standard Albanian. Note that in the next section, Table 2 shows *gërshërë* ‘scissors’ as one of the words used as stimuli for the vowel nasality feature. From the orthography, it may not be obvious that the Gheg form of this word is /gərʃãn/. This is because of a series of sound changes that took place in Tosk, on which the orthography is based. Vowel nasality likely developed historically because of coarticulation with a nasal consonant. Tosk subsequently lost this feature, in addition to undergoing rhotacism, another sound change whereby intervocalic /n/ became a rhotic consonant. The modern form of the word in Tosk is thus /gərʃərə/ (pronounced [gərʃərə] or [gərʃər]) where the last consonant is a nasal-turned-rhotic and the preceding vowel was denasalized.

Speech material

The verbal guise technique was used for this study (Cooper, 1975; Garrett, 2010:53–87). The stimuli were non-modified, isolated words produced by 48 native speakers of Albanian living in Albania, whose demographic characteristics are shown in Table 1.

Table 1. Basic demographic characteristics of the 48 speakers who produced the stimuli

Dialect	Number of speakers	Mean age	Sex
Gheg	26	52	19 F, 7 M
Tosk	22	42	14 F, 8 M

Twenty-one of these speakers participated in a picture-naming task in which isolated target words were produced (e.g., *spinaq* ‘spinach’), while the remaining 27 participated in a reading task where the same target words were produced in initial and final positions of carrier sentences (e.g., *spinaq thoni spinaq* ‘spinach, say spinach’). Speakers were digitally recorded (44,100 Hz, 16 bits) using *Speech Recorder* (Draxler & Jansch, 2004), a Tascam US-2x2, and a Beyerdynamic TG H54c head-mounted microphone, in quiet locations in the municipalities of Ballaban, Përmet (Tosk area), Bërzhitë, and Tirana (Gheg area).

For this study, 64 stimuli were used (i.e., 32 tokens per dialect). For each dialect, two repetitions by different speakers of the 16 words shown in Table 2 were selected. In order to select 48 tokens for rounding of /a/, monophthongization, and vowel nasality, we first asked three native speakers of Albanian (who were neither speakers nor listeners in the experiment) to listen to a longer set of stimuli and report whether the relevant features had been produced. A selection was then made among the tokens that the three consultants unanimously agreed upon. To select 16 tokens for contrastive vowel length, we relied on acoustic measurements of vowel duration. We first computed the mean and standard deviation per word and per dialect, then selected tokens which had vowels with a duration of approximately one standard deviation above the Gheg mean for Gheg, and one standard deviation below the Tosk mean for Tosk. This resulted in tokens with a mean vowel duration of 227 ms for Gheg and 102 ms for Tosk. Two stimuli were also selected to serve as practice trials at the beginning of the experiment:

the word *llokum* ‘lokum, Turkish delight’ from a female Gheg speaker and the word *kafe* ‘coffee’ from a male Tosk speaker. *Praat* (Boersma & Weenink, 2023) was used to manually isolate and chop the selected words from longer sound files, to equalize their intensity at 60 dB, and for silence-padding (200 ms silences at each end).

Table 2. The 16 words of the experiment and corresponding feature; 2 tokens per word per dialect were used, for a total of 64 stimuli

Feature	Tosk variant	Gheg variant	n words	Albanian stimulus words and English gloss			
Rounding of /a/	[a]	[ɔ]	4	domate ‘tomatoes’	i madh ‘large’	mal ‘mountain’	spinaq ‘spinach’
Contrastive vowel length	/V/	/V:/	4	bukë ‘bread’	kalë ‘horse’	mi ‘mouse’	raki ‘raki’
Monophthongization	/VV/	/V/	4	diell ‘sun’	duar ‘hands’	fyell ‘flute’	pastruese ‘cleaner’
Vowel nasality	/V/	/Ṽ/	4	gërshërë ‘scissors’	hënë ‘moon’	nëntë ‘nine’	shkëmb ‘rock’

Procedure

The experiment lasted approximately 25 minutes and was run online using *PsyToolkit* (Stoet, 2010, 2017), with the entire interface and instructions in Albanian. After agreeing to informed consent and confirming they lived in Albania, the participants performed one listening task featuring the 2 practice trials and 64 test stimuli. They were instructed to use scales and sliders to give their spontaneous opinion of the *voices* they heard. This was done to avoid that participants felt they were judging people, which according to our three consultants, is considered rude in Albania and could make potential participants reluctant to take part in the experiment. They were given as an example a scale featuring the adjectives *young* (*i ri*) and *old* (*i moshuar*) at each pole, and were told to move the slider at the *young* pole if the voice sounded young, at the *old* pole if the voice sounded old, and toward the *young* pole but closer to the middle of the scale if the voice sounded somewhat young. No audio accompanied this example.

After the two practice trials, the order in which the stimuli were presented was randomized. Each stimulus could be replayed an unlimited number of times and had to be rated on five VASs using sliders that could be positioned anywhere between two adjectives corresponding to the poles of each scale, as shown in Figure 1. Although the scales appeared continuous in the user interface, they were made up of 101 equally spaced discrete steps (i.e., from 0 to 1 inclusively, in .01 increments). The sliders were initially centered at the .5 mark, and all had to be “activated” by a mouse click or finger tap for the participants to proceed to the next stimulus. As shown in Figure 1 and Table 3, one scale was for dialect identification, and two pertained to status of and solidarity with the speaker respectively.³ The remaining two scales were fillers and asked about perceived age (used for the instructions) and gender, but these are not analyzed here. The order of the scales was randomized across stimuli and participants, but the order of the poles was the same throughout the experiment for everyone. Because there was a

risk that out-of-context isolated words would be misunderstood or incorrectly mapped (e.g., Dragojevic & Giles, 2016), pictures always accompanied the audio stimuli.

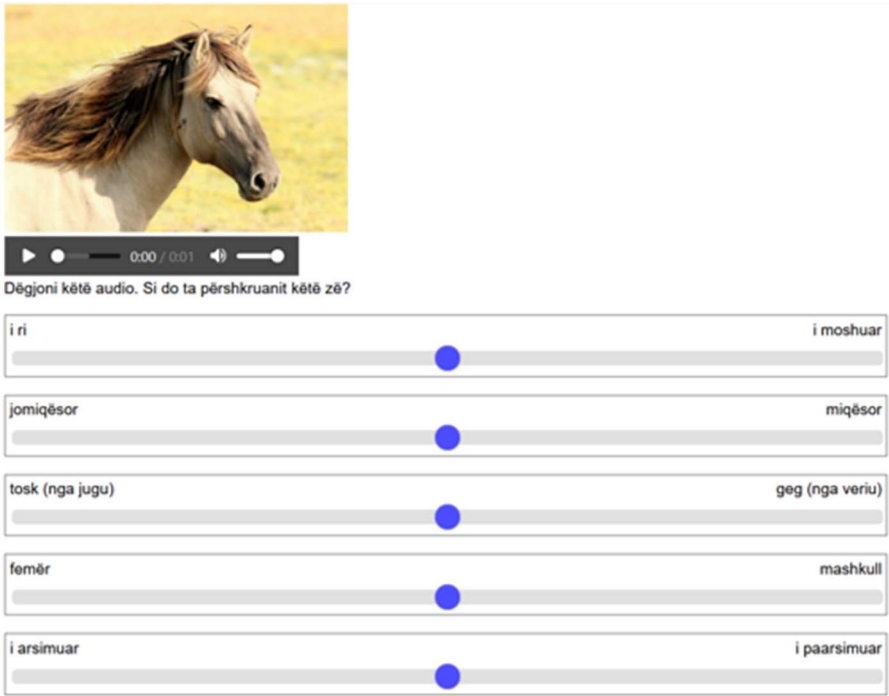


Figure 1. Example of interface of the experiment (desktop version). This picture accompanied stimulus word *kalë* 'horse'. The pictures used were either under *Pixabay* license or created by the authors in Microsoft Office.

Table 3. Pairs of adjectives used as poles for the five scales

Scale	Albanian pole adjectives	English gloss
Dialect identification	tosk (nga jugu)–geg (nga veriu)	Tosk (south)–Gheg (north)
Status	i arsimuar–i paarsimuar	educated–uneducated
Solidarity	jomiqësor–miqësor	unfriendly–friendly
Age (filler)	i ri–i moshuar	young–old
Gender (filler)	femër–mashkull	female–male

After the listening task, the participants provided the following information about themselves: age, sex, place of birth, place of residence, other places where they had lived and for how long, and origin of their parents. Upon completion of the whole experiment, the participants were compensated with a €5 voucher to be used in a bookstore in Tirana.

Listeners

Listeners were recruited among the second author's network, then by snowball sampling. One hundred thirty-three listeners completed the task. We excluded responses from six listeners who completed it much faster than the others and whose response patterns suggested they were not attending to the task, and two listeners whose dialect background could not be identified. The remaining 125 participants considered in this study were 87 female and 38 male aged 18–52 years old (mean 27). In order to determine whether they had a Tosk or a Gheg background, we used their residential history and their parents' origin. One hundred four participants were still living in the same dialect area where they were born, and were therefore categorized as Gheg or Tosk based on this geographical information. For the remaining 21 participants who had moved across dialect areas, we considered their dialect to be that of the area where they had spent the majority of their lives before 18 years old, and that of their parents. There were no ambiguous cases other than the two exclusions mentioned above. This resulted in 34 listeners with a Tosk background and 91 with a Gheg background.

We believe that the sex and dialect imbalances, as well as the relatively young age of our sample, are largely due to the sampling technique and to the online format especially, which failed to reach certain segments of the population. However, these imbalances are not a major issue with the statistical approach explained in the next section.

Statistical analyses

The aim of the statistical analyses was essentially to estimate the typical (or average) ratings that listeners gave depending on the scale, feature, and dialect represented. The choice of an appropriate statistical model architecture was guided by the distributional properties of the collected data. To this effect, [Figure 2](#) illustrates the distribution of ratings pooled across the entire data set. Though defined on 101 discrete values, the rating distribution appears smooth enough to be modeled as continuous. It is bounded between 0 and 1, and exhibits two major peaks at the extremes, as well as a minor one at .5. We thus opted for a zero-one inflated beta (ZOIB) model, currently considered the state of the art for modeling continuous bounded responses with peaks at both ends (e.g., Bendixen & Purzycki, 2023; de Hoop et al., 2023; Liu & Eugenio, 2018). A ZOIB model is a combination of four sub-models, which account for different subsets of the response. A logistic model (*zoi*) predicts whether the response equals one of the two extreme values, standardized as 0 and 1, or lies inside the continuous interval (0, 1). The subset of non-extreme responses is modeled by beta regression, which entails the estimation of two parameters of the beta distribution, namely *mean* and precision (*phi*). Finally, the subset of extreme responses is modeled by another logistic regression (*coi*) predicting whether the extreme is 0 or 1.

The estimated regression models accommodated the presence of groups of correlated responses (by listener) by introducing so-called random effects (e.g., Kingston, Baayen, & Clopper, 2012). We opted for a Bayesian (rather than frequentist) approach because the estimation of mixed-effect generalized linear models involving hundreds of parameters (as is the case here) tends to be less affected by convergence issues when carried out with a Bayesian toolkit, thanks to the application of Markov chain

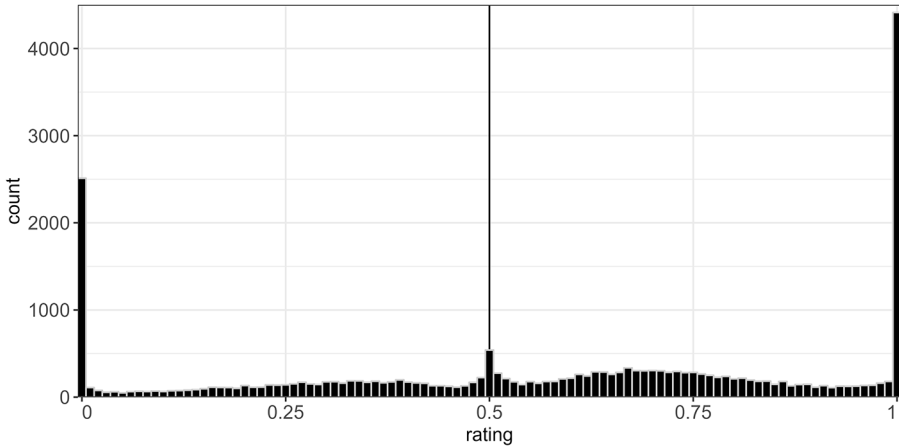


Figure 2. Distribution of ratings on the three scales for the four features. Each .01 increment from 0 to 1 is represented by a bin, with higher bins showing higher rating counts.

Monte Carlo (MCMC) (see Winter, Fischer, Scheepers, & Myachykov, 2023 for similar motivations; for recent work featuring Bayesian modeling of speech perception and production data, see Cole, Steffman, Shattuck-Hufnagel, & Tilsen, 2023 and Roettger, Franke, & Cole, 2021). Even so, as a measure of compromise between complexity and feasibility, we decided to model the responses on each scale separately.

Three Bayesian mixed-effect ZOIB regression models were fitted, one per scale of interest, with the listeners' ratings of the stimuli as response variable.⁴ All ZOIB sub-models, except for the one estimating beta precision, had the same predictor composition. Fixed predictors were *stimulus dialect* (Tosk or Ghég), *listeners' dialect* (Tosk or Ghég), and *feature* (four levels), together with all their two- and three-way interactions. Random intercepts for listener, as well as random slopes for stimulus dialect, feature, and their interaction over listener were introduced. Beta precision sub-models contained only a fixed intercept and a random intercept for listener, similarly to de Hoop et al. (2023).

Models were fit using the *zero_one_inflated_beta* setting from the *brms* package (Bürkner, 2017) in *R* (version 4.3.x; R Core Team, 2023). The models were run with four MCMC chains, 8000 iterations, and 2000 warm-up samples. In all cases, model convergence was achieved ($\hat{R} = 1$, in some cases $\hat{R} = 1.01$; see Supplementary Materials). Priors were left to their default settings. Posterior *z*-scores and shrinkage (Betancourt, 2018) indicated that the priors did not dominate the posterior distributions (see Supplementary Materials), while a few parameters of the *coi* sub-models were poorly identified, which might have to do with the extreme values not being a high proportion of the responses. The posterior predictive check plots show that the models fitted the data reasonably well, save for cases where the data exhibited mild bimodality trends. The mean Bayesian pseudo- R^2 were .406, .241, and .291 for the models predicting dialect identification, solidarity, and status ratings respectively, which is not very high (see de Hoop et al., 2023 for similarly modest performance on responses obtained from

sliders and modeled by ZOIB). As illustrated in the supplementary materials, this is likely due to the tendency of listeners to use the center of VASs, which cannot be captured by any ZOIB sub-model, and by the great variability in response strategies across participants, which can only be partly modeled by introducing random intercepts and random slopes over listener (see de Hoop et al., 2023 for similar findings). In order to assess the impact of these issues, we fitted for each scale a companion Bayesian logistic regression model where continuous ratings were reduced to a binary alternative between 0 and 1 (see Supplementary Materials). These simplified models corroborated the results from the ZOIB ones in showing the same qualitative trends in all cases; the results presented in this article are thus those from the ZOIB.

Using the *emmeans* package (Lenth, 2023), conditional posterior distributions of the mean ratings⁵ were calculated for each combination of the fixed predictors, for a total of 16 combinations per model (2 stimulus dialects \times 2 listeners' dialects \times 4 features), from which medians (M), and 66% and 95% credible intervals were generated. These were plotted as in Figure 3, in which computed medians are represented by vertical ticks, 66% credible intervals by thick horizontal lines, and 95% credible intervals by thin horizontal lines. To give an example of how to interpret the results in this figure, shaded box A shows that the rating of Gheg stimuli by Gheg listeners for rounding of /a/ is predicted to be between .336 and .371 with 95% probability (i.e., the extremes of the thin horizontal dark gray line on the left-hand side of box A). As this interval lies entirely on the left of the .5 mark and by a fairly large margin, we can argue that Gheg listeners rated this type of stimuli as sounding more uneducated than educated.

Conditional posterior distributions were also computed for expected differences between predictor values in three ways, corresponding to boxes A, B, and C in Figure 3. First, *inter-stimulus* differences were computed between ratings of Tosk and Gheg stimuli on a given scale, within feature and within listener dialect. This is illustrated by shaded box A, where the inter-stimulus difference is that between status ratings attributed to Tosk and Gheg stimuli by Gheg listeners for rounding of /a/. The inter-stimulus difference is defined as the rating attributed to Tosk minus the one attributed to Gheg, or formally: $rating(stimulus: Tosk) - rating(stimulus: Gheg)$. If it is positive, Tosk stimuli were rated higher than Gheg. The larger the inter-stimulus difference (whichever the sign), the more distinctly Tosk and Gheg stimuli were rated. In box A, it is about +.32.

Second, *inter-listener* differences were computed between ratings of Tosk and Gheg stimuli on a given scale, *within* feature but *between* listener dialects. This is a difference-of-differences like that illustrated in dotted box B of Figure 3, where first, two inter-stimulus differences were computed, then these were subtracted from each other, formally: $inter-stimulus(listener: Tosk) - inter-stimulus(listener: Gheg)$. A small inter-listener difference indicates that Tosk and Gheg listeners expressed a similar separation between the ratings of Tosk and Gheg stimuli. In box B, the inter-stimulus difference for Tosk listeners (right) is about +.05 and for Gheg listeners (left) +.09. Therefore, the inter-listener difference is about $-.04$ ($.05 - .09 = -.04$).

Third, *inter-feature* differences were computed between ratings of Tosk and Gheg stimuli on a given scale, *across* features but *within* listener dialect. This is a difference-of-differences like that illustrated in dashed box C of Figure 3, formally: $inter-stimulus(feature: monophthongization) - inter-stimulus(feature: vowel nasality)$.

After computation of inter-feature differences for every pair of features for a given group of listeners, the features can be ranked; those ranking highest can be considered to have more “discriminatory power” than low ranking ones; whenever the 95% credible interval of an inter-feature difference overlaps with zero, the two features are assigned equal ranking. In box C, vowel nasality ranks higher than monophthongization, as the inter-stimulus difference for the former is about .09 larger than for the latter.

Due to space constraints, only a selection of results based on differences between predictor values will be reported in the text (see Supplementary Materials for the full set).

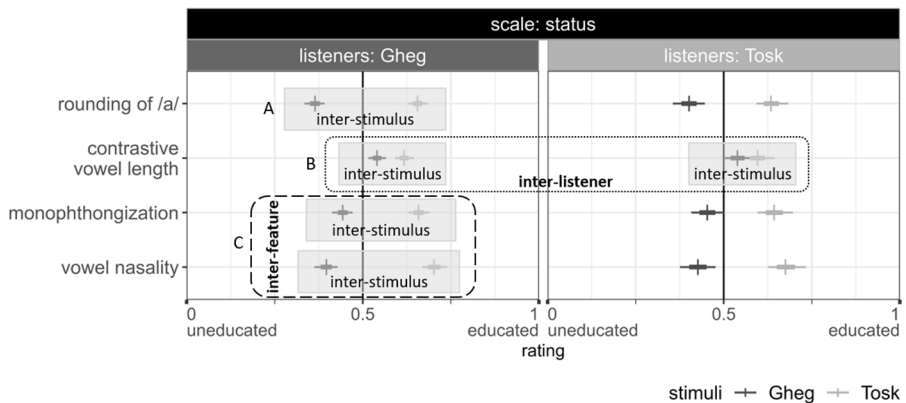


Figure 3. Medians and credible intervals for the conditional mean ratings on the status scale for the four features, separated by dialect of the speakers who produced the stimuli (tick shade) and dialect of the listeners (panels). Boxes A to C illustrate the three types of comparisons carried out. An unannotated version of this figure reappears as Figure 7.

Results

In each subsection, first raw rating distributions are described, then results from the ZOIB models are reported.

Dialect identification

The distributions of the listeners’ raw ratings on the dialect identification scale are displayed in Figure 4. In the four panels of Figure 4 (as well as Figures 6 and 8), the dark gray curves correspond to rating distributions for Gheg stimuli, the pale gray curves to distributions for Tosk stimuli, the solid curves to distributions for Gheg listeners, and the dashed curves to distributions for Tosk listeners. The center of the scales is indicated by a black vertical line.

Figure 4 suggests that Tosk stimuli were accurately identified as Tosk for the four features, with little difference between ratings produced by Tosk and Gheg listeners. Gheg stimuli were accurately identified when they featured rounding of /a/, monophthongization, and vowel nasality. The rightmost panel shows a more extensive use of

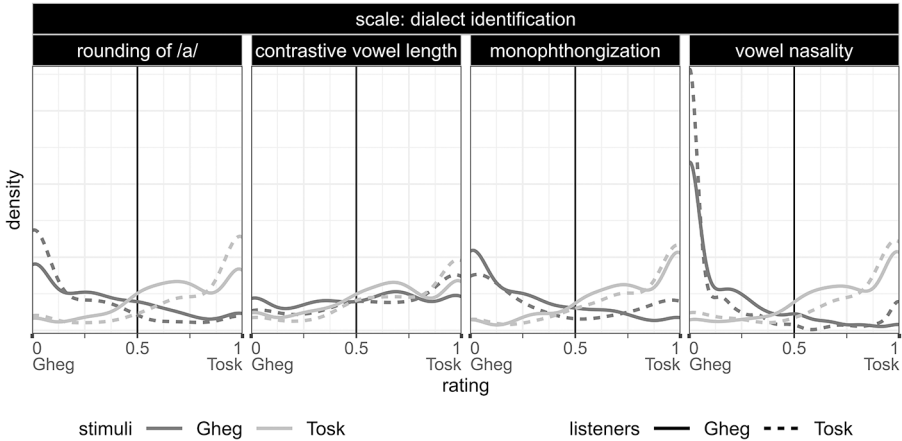


Figure 4. Density plots of the listeners' raw ratings on the dialect identification scale for the four features, depending on whether the stimuli were produced by Gheg or Tosk speakers (shade), and whether the listeners were of Gheg or Tosk background (line type).

the Gheg pole of the scale for vowel nasality than for other features, as indicated by high peaks at the Gheg extremity. For contrastive vowel length, however, there was a tendency for Gheg stimuli to be identified as Tosk. Gheg listeners were particularly undecided, as suggested by the absence of clear high peaks and low dips in their data distribution.

The data were analyzed using the ZOIB regression models described before. In Figure 5 (as well as Figures 7 and 9), the posterior values are reported separately for Tosk and Gheg stimuli, Tosk and Gheg listeners, and the four features. The vertical ticks correspond to the medians (M), the thick horizontal bars to the 66% credible intervals, and the thin horizontal bars to the 95% credible intervals of the conditional mean ratings. The center of the scale (value .5) is indicated by a black vertical line.

Consistent with the raw rating distributions, the model predictions suggest ratings toward the Tosk pole of the scale for Tosk stimuli (i.e., above .5). The opposite is true for Gheg stimuli, except when these featured contrastive vowel length, in which case their ratings were above .5, and the inter-stimulus differences were negligible (Gheg listeners: M , .10; Tosk listeners: M , .07; compared to, respectively, M , .53 and M , .55 for vowel nasality, for instance). The inter-listener comparisons suggest very similar rating behaviors by Gheg and Tosk listeners, except for a tendency for Tosk listeners to rate Gheg and Tosk stimuli more distinctly when these featured rounding of /a/ than the Gheg listeners did (M , .18). Consequently, the inter-feature comparisons suggest the following ordering of features in terms of discriminatory power: for Gheg listeners, vowel nasality > monophthongization > rounding of /a/ > contrastive vowel length; for Tosk listeners, vowel nasality > rounding of /a/ > monophthongization > contrastive vowel length. Stimuli featuring vowel nasality thus triggered the most distinct ratings in both groups of listeners and contrastive vowel length the least distinct ratings.

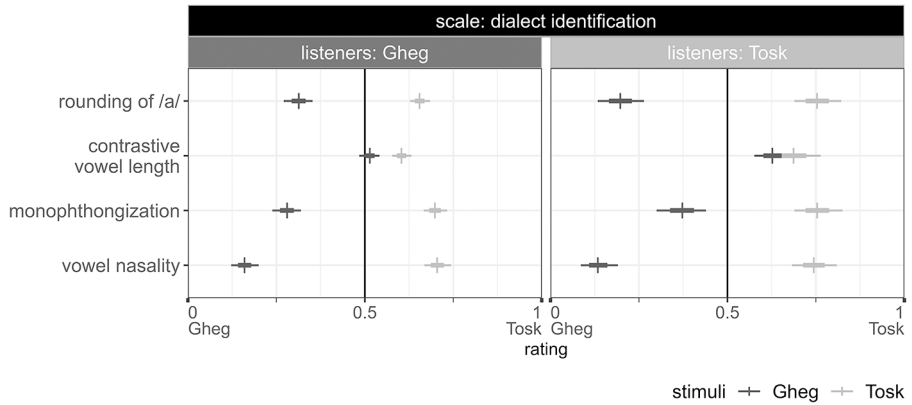


Figure 5. Medians and credible intervals for the conditional mean ratings on the dialect identification scale for the four features, separated by dialect of the speakers who produced the stimuli (tick shade) and dialect of the listeners (panels).

Status

Figure 6 shows the distributions of the listeners’ raw ratings on the status scale. The panels corresponding to rounding of /a/, monophthongization and vowel nasality indicate that for these features, Ghgeg stimuli were rated more toward the uneducated end of the scale, and Tosk toward the educated end. For rounding of /a/, Tosk listeners rating Tosk stimuli used more the center of the scale than Ghgeg listeners rating Tosk stimuli. Tosk stimuli featuring vowel nasality received the most polarized ratings. The panel displaying data for contrastive vowel length does not show obvious differences in how educated Ghgeg and Tosk stimuli were rated. The results for status pattern with those for dialect identification presented in the previous section: stimuli identified as Tosk were rated as more educated, while stimuli identified as Ghgeg were rated as less educated.

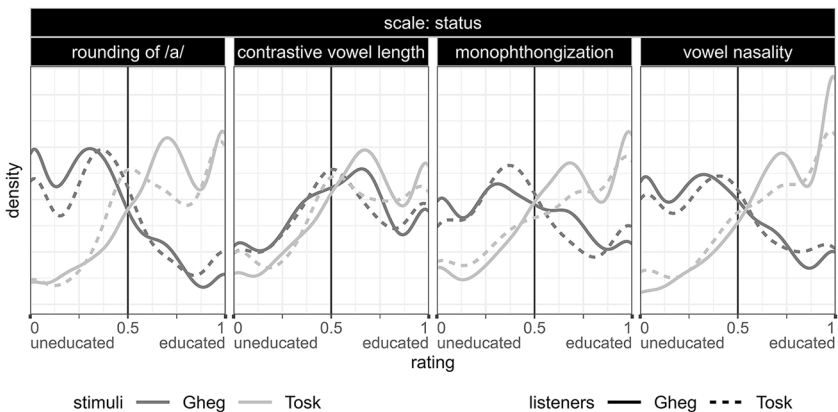


Figure 6. Density plots of the listeners’ raw ratings on the status scale for the four features, depending on whether the stimuli were produced by Ghgeg or Tosk speakers (shade), and whether the listeners were of Ghgeg or Tosk background (line type).

The model predictions presented in Figure 7 suggest ratings above .5 for Tosk stimuli, that is, more toward the *educated* end of the scale, by both listener groups and for all features. The Gheg stimuli were rated more toward the *uneducated* end, with estimates below .5, for three out of four features. The Gheg stimuli for contrastive vowel length were instead rated more similarly to the Tosk ones, with values above .5, by both groups of listeners. The inter-stimulus difference is indeed negligible for contrastive vowel length (Gheg listeners: M , .08; Tosk listeners: M , .05). The inter-listener comparisons suggest slightly more distinct ratings from Gheg than Tosk listeners (rounding of /a/: M , -.07; contrastive vowel length: M , -.03; monophthongization: M , -.04; vowel nasality: M , -.05), but the inter-feature comparisons reveal the same ranking of features in both groups, with rounding of /a/ and vowel nasality equally ranked as having triggered the most distinct ratings, followed by monophthongization, and by contrastive vowel length.

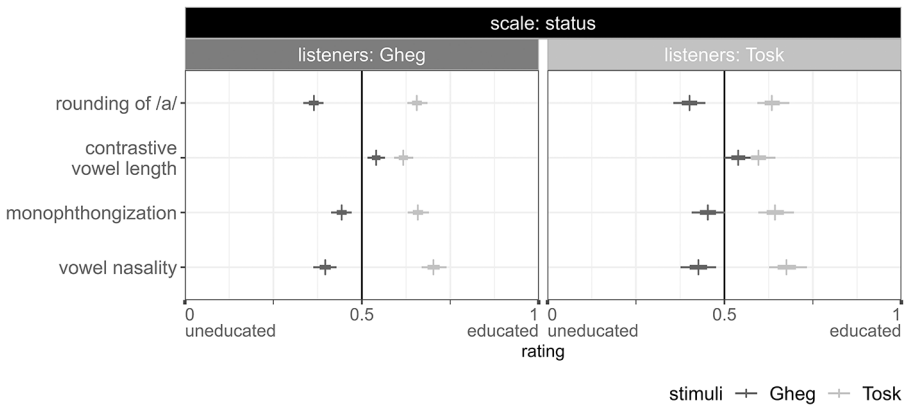


Figure 7. Medians and credible intervals for the conditional mean ratings on the status scale for the four features, separated by dialect of the speakers who produced the stimuli (tick shade) and dialect of the listeners (panels).

Solidarity

Figure 8 suggests that both Gheg and Tosk stimuli were generally rated as more friendly than unfriendly for the four features. This response pattern is quite distinct from those obtained for the dialect identification and status scales. However, in a similar manner to the previous scales, few differences emerge from the solidarity ratings produced by Gheg and Tosk listeners.

All model predictions in Figure 9 lie above .5, toward the *friendly* pole of the scale, irrespective of the feature, listener group, and stimulus dialect. Inter-stimulus differences are minimal, though they are notably reversed for contrastive vowel length, for which Gheg stimuli were rated marginally higher than Tosk stimuli (Tosk listeners: M , -.03; Gheg listeners: M , -.03; inter-listener difference: .002). The inter-listener comparisons show that monophthongization triggered slightly more distinct ratings among Tosk listeners (M , .04). The inter-feature comparisons suggest the following ranking

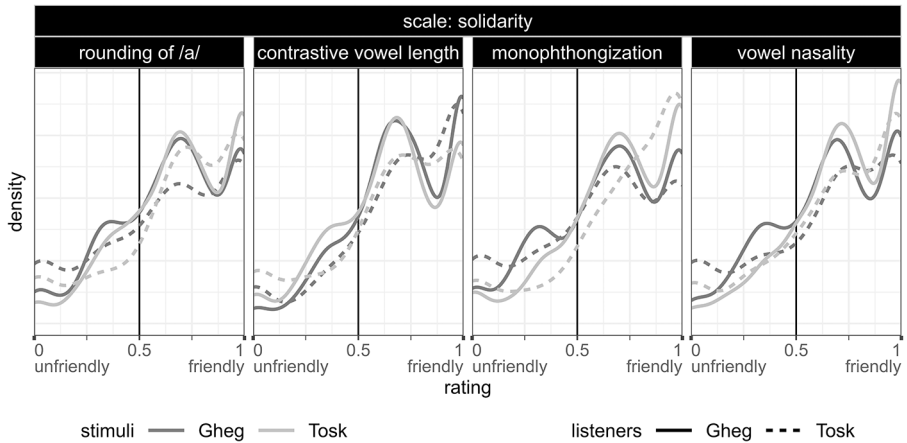


Figure 8. Density plots of the listeners’ raw ratings on the solidarity scale for the four features, depending on whether the stimuli were produced by Ghgeg or Tosk speakers (shade), and whether the listeners were of Ghgeg or Tosk background (line type).

for Ghgeg listeners: rounding of /a/, monophthongization, and vowel nasality ranked equally, followed by contrastive vowel length, which triggered the least distinct of not very distinct ratings overall. For Tosk listeners, monophthongization ranked highest, followed by rounding of /a/ and vowel nasality which ranked equally, and finally by contrastive vowel length like for Ghgeg listeners.

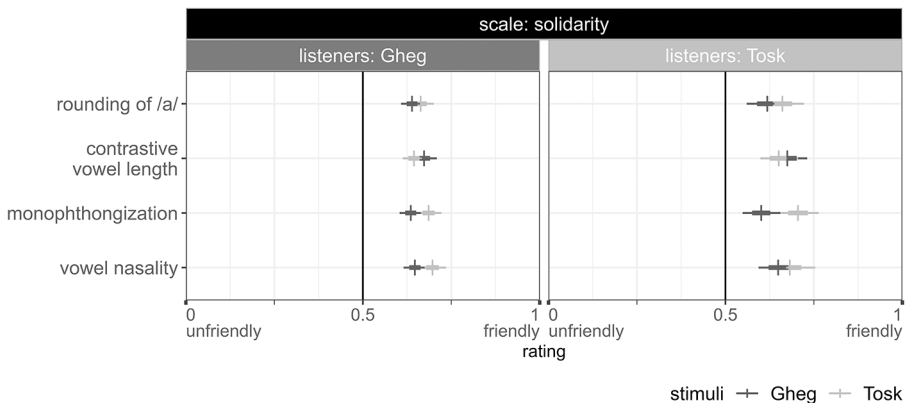


Figure 9. Medians and credible intervals for the conditional mean ratings on the solidarity scale for the four features, separated by dialect of the speakers who produced the stimuli (tick shade) and dialect of the listeners (panels).

Discussion

The aim of this study was to document Albanians’ responses on dialect identification, status, and solidarity scales for four features differing between Ghgeg and Tosk, motivated by recent work which found some features of Tirana Ghgeg were changing

while others were not, a picture which language-internal factors could not entirely account for.

A dialect identification scale was integrated into the design to verify whether the attitudes expressed by the participants could be interpreted as being directed at dialects. Overall, the results suggested that this interpretation held for rounding of /a/, monophthongization, and vowel nasality. For these three features, listeners were fairly accurate at identifying dialects using the limited information contained in isolated words. Nasal vowels appeared to be a particularly strong cue for the identification of Gheg stimuli, as suggested by the listeners' polarized responses for the Gheg variant of vowel nasality (Figure 4) and the finding that this feature triggered the most distinct ratings between Tosk and Gheg stimuli. Combined with the comments collected by Morgan (2015), these trends point to vowel nasality being a *stereotyped* feature of Gheg, in Labov's (1972:314) terms. On the other hand, the fairly robust reaction also measured for rounding of /a/ and monophthongization suggests that these features are at least *markers*.

The results on the status scale for rounding of /a/, monophthongization, and vowel nasality showed that Tosk stimuli were rated as sounding generally more educated than uneducated, while the opposite was found for Gheg stimuli. As explained before, the advent and promotion by the Albanian state over the past 50 years of a standard language that was much more similar to Tosk than Gheg may have contributed to Albanians forming such an opinion. Bugge (2018:327) proposed that "the codification of a standard spoken language and the establishment of a standard language ideology [were] essential to the establishment of status hierarchies of spoken varieties." This would account for why, in linguistic settings lacking a standard like the Faroe Islands (Bugge, 2018) or Western Norway (Anderson & Bugge, 2015), dialects' status is higher than in settings with an established standard, like Denmark (Kristiansen, 2009), Lithuania (Vaicekauskienė, 2019), or Albania as shown in this study.

The results relative to dialect identification and status were different for the fourth feature, contrastive vowel length. Dialect identification, in particular, was much less distinct and accurate. As detailed in the Methods section, stimulus selection was done differently for words featuring contrastive vowel length—that is, based on measured acoustic characteristics, compared to perceptual assessment for the other features. We do not think that the acoustic-based selection per se is the cause of lower accuracy. First, with mean vowel durations of 227 ms for Gheg stimuli and 102 ms for Tosk stimuli, the contrast was well above the threshold of discriminability by human listeners (e.g., Casini, Burle, & Nguyen, 2009; Chiu, Rakusen, & Mattys, 2019; Klatt, 1976), making the task unquestionably feasible from a psychoacoustic point of view. In addition, by selecting stimuli with a duration within the range of one standard deviation from the mean, these fell within a plausible duration spectrum for short and long Albanian vowels (see Lehnert-LeHouillier, 2010 on language-specific thresholds). The difference in duration is also similar to that found in other languages with contrastive vowel length (e.g., Paschen, Fuchs, & Seifart, 2022). The reason for using a different stimulus selection technique for this feature was because the consultants who helped selecting the other ones found the task too difficult, even considering that selecting other features was not particularly easy (e.g., only 33% inter-consultant agreement for monophthongization). If anything, this was a first insight that contrastive vowel length would receive

a different evaluation from the other features, and justified testing it on a larger sample. This also suggests that contrastive vowel length could hold the status of an *indicator*.

Ratings on the status scale for contrastive vowel length indicated that both Gheg and Tosk stimuli were perceived as more educated than uneducated, in contrast with the previous features. The high status attributed to Gheg long vowels might explain why this feature is preserved in Tirana Gheg while others with lower status have been found to change toward more prestigious variants. However, we also have to consider that for this feature, listeners were not successful at identifying Gheg stimuli as being Gheg. It thus seems to be less about Gheg long vowels benefiting from a high status, than Gheg vowels not being identified as such. For Tosk listeners, this poorer identification performance could perhaps be explained by a lack of familiarity with Gheg (e.g., Adank, Evans, Stuart-Smith, & Scott, 2009; Clopper & Pisoni, 2004a) given the marginal place of this dialect in mainstream media due to standardization. However, this explanation does not hold for Gheg listeners, who were obviously familiar with their own dialect, in addition to being substantially exposed to standard Albanian and/or Tosk through media and education. Chen, Rattanasone, Cox, and Demuth (2017) presented some evidence that familiarity with another dialect increased phonological flexibility and tolerance to phonetic variation in vowel length in Australian English listeners, and proposed, at least for their bidialectal subjects, that the vowel length feature was not well specified in the lexicon. Such an explanation for the relatively low performance of Gheg listeners at identifying their own dialect based on long vowels remains speculative at this stage, but may point to some difference in the cognitive representation and processing of length compared to other vowel features.

Gheg and Tosk stimuli corresponding to the four features were rated as more friendly than unfriendly. This may not be surprising for Gheg, given that low status varieties tend to trigger high solidarity. The finding that Tosk was also judged as friendly by both groups of listeners, as evidenced in this study and in Dickerson (2021), could point to a sense of national solidarity fostered by various potential factors. One of these could be the great emphasis put for half a century or so by the former People's Socialist Republic of Albania on unifying Albanians into one people (Kostallari, 1970:26). Albanians' mutual tolerance in other cultural spheres such as religion has also been cited as exceptional within the European context (Kruja, 2020; UNDP, 2018). In addition, the idea of Albanians standing united against other ethnic groups or nationalities formed part of the backdrop against which a national identity arose from 1912, after Ottoman withdrawal from the Balkans (e.g., Bego, 2020; Xhudo, 1995). Whether or not these factors have contributed to the participants perceiving each other as friendly, Tosk is not a unique case of high status, high solidarity variety. For Singapore Standard English, which was highly rated on both dimensions, Cavallaro and Chin (2009:155) suggested a possible role of government-sponsored campaigns overtly promoting the standard.

Some differences emerged between responses provided by Tosk and Gheg listeners. When responding on the dialect identification scale, Tosk listeners rated Gheg and Tosk stimuli more distinctly when these featured rounding of /a/ than the Gheg listeners did, while the opposite was found for monophthongization. On the solidarity scale, Tosk listeners produced more distinct ratings of Gheg and Tosk stimuli featuring monophthongization than the Gheg listeners did. While these differences have influenced the

scale-contextual ranking of features obtained from inter-feature comparisons, their magnitude was so small that we are not confident these trends could be replicated. Broadly speaking, our results are thus compatible with the recurrent observation that members of a speech community are consistent in their judgments.

Overall, our results show differences across features which suggest a connection between attitudes and dialect stability and change: the two features previously found to change in Gheg, rounding of /a/ and monophthongization, were attributed a low status and were clearly identified as Gheg; whereas the stable feature, contrastive vowel length, was not convincingly identified as Gheg and was attributed a high status. In the Albanian context, attitudes thus seem to fill an explanatory gap left by only considering language-internal factors. This leads us to argue that attitudes are a language-external factor worth considering alongside language-internal ones in studies of language change, as low status features are more prone to be replaced. Models which seek to explain sound change, and in particular the *actuation* problem, might always miss a piece of the puzzle if attitudes are not considered (Pinget, 2015).

Labov (2006:203) introduced a distinction between linguistic change from above and from below, mainly based on speakers' awareness. While changes from below pass under the speech community's radar, changes from above benefit from full awareness, and usually involve prestigious variants whose social distribution is reshuffled. Evidence was found in this study that listeners were aware of co-existing variants, and that they judged more favorably the Tosk ones. Variants of changing features are currently redistributed within Albanian society, with Gheg speakers living in Tirana adopting the Tosk/standard ones (Riverin-Coutlée et al., 2022). The dynamic of these changes is characterized by an urban-rural divide, as Tirana Gheg was found to be at a more advanced stage than Gheg spoken in a rural area nearby (Riverin-Coutlée et al., 2022). In Morgan's (2015:49) study, there were some indications that the interviewed participants were attuned to this urban-rural divide: they hypothesized that vowel nasality and monophthongization, which they occasionally encountered in Tirana, had been "borrowed" from more remote areas of Gheg and added to Tirana speech which was otherwise qualified as standard. In reality, Tirana is located in a Gheg-speaking area, and has always been described as comprising these features, but because the participants' attitudes toward Gheg conflicted with the urbanity of Gheg speakers living in Tirana, an excuse was made up to justify the presence of certain dialect features in their speech. All this points to changes from above, with a high degree of awareness of the changing features, and the speech community conceptualizing feature (re)distribution in a way that is somewhat aligned with the documented course of the changes, that is, characterized by an urban-rural divide.

There are limitations to this study which hinder the generalizability of our findings to all Albanians. As mentioned before, there are segments of the population which we could not reach with an online experiment, as reflected for instance in the relatively young age of the sample. We cannot exclude that the results would have been different had we collected responses from older participants with limited capacity to access the internet. Another limitation has to do with the response format. VASs allow expressing nuanced distinctions in ratings which are impossible to capture with, say, Likert scales, where the pre-determined set of labeled alternatives is small. Likert scales are affected by a number of limitations and biases, like the "central tendency" bias (Sims, 2002:94),

which VASs eliminate or at least mitigate (Llamas & Watt, 2014). However, the less constrained interface of VAS opens the door to more variability, whose origin may or may not be related to the phenomenon under investigation. In this study, a great deal of response variation can be ascribed to inter-listener differences in their interaction with VAS, where some individuals tend to remain close to the midpoint, others to use the whole range, others to use only the extremes. While this appears to be quite common (e.g., de Hoop et al., 2023:Figure 3; Kim, Clayards, & Kong, 2020:Figure 5) and not exclusive to language-related phenomena (e.g., Wentzky & Summers, 2020 on rating videos of factory tasks), it is still not established whether individual approaches to VAS reflect aspects of relevance for a given study, or general cognitive strategies in the interaction with the rating interface (van Osch & Stiggelbout, 2005).

High inter-listener variability posed a challenge to statistical modeling, which we tackled in two steps. First, we fitted Bayesian ZOIB regression models, which to our knowledge are currently among the most sophisticated ones for the analysis of VAS data (see Liu & Eugenio, 2018 for a review). Second, since the fitted models did not score very high in terms of explained variance (similarly to de Hoop et al., 2023), we compared the results to a companion set of Bayesian logistic regressions applied to a binary version of the response data, which confirmed all the general qualitative trends obtained with ZOIB models. We argue that the difficulties we encountered in the modeling may reveal an absence of statistical instruments targeted to data that exhibit strongly characterized latent profiles, which may not be effectively factored out by the classic hierarchical structure provided by mixed-effect regression (ZOIB or other).

Conclusion

This study showed that understanding dialect and language change benefits from considering attitudes. It also provided further evidence that listeners are sensitive to government-promulgated language ideology. As VASs are increasingly used in various fields, the study identified critical difficulties in handling VAS data and areas of improvement in terms of knowledge and modeling. Finally, due to our methodological choices, a question that remains open is whether the evaluation of a feature may influence that of another (Campbell-Kibler, 2011; Montgomery & Moore, 2018; Pharaon, *in press*; Watson & Clark, 2015). For example, could phonetic features be overlooked when a syntactic feature is present? What the study does show, however, is that a single word is clearly enough to trigger a measurable reaction (Scharinger, Mohahan, & Idsardi, 2011) and for people to make judgments which may then play a role in language change.

Supplementary material. The supplementary materials, as well as the data and code necessary to replicate the analysis, are accessible here: <https://osf.io/cp4at/>.

Acknowledgements. Data collection was supported by an Alexander von Humboldt fellowship (author E.K.) and by the project *InterAccent* funded by the European Research Council (advanced grant 742289, 2018–2023, awarded to Jonathan Harrington). Many thanks go to Jeta Alla, Dijon Ismaili, and Sonja Krasniqi for their insights, and to the editorial team and reviewers for their feedback.

Competing interests. The authors declare none.

Notes

1. Some authors make a link between stereotypes and the notion of salience (e.g., Montgomery & Moore, 2018).
2. These are not the terms used in Morgan's (2015) study, who instead bundled together various phenomena under the labels "Tirana O" and "shortening," which include vowel nasality and monophthongization respectively.
3. These specific pairs of adjectives were selected from the longer list in Zahn and Hopper (1985) because they were easily translatable into Albanian and because we and our three consultants judged them culturally meaningful and appropriate.
4. For ease of exposition, the poles of two scales were reversed compared to the original experiment: 0 now corresponds to *Gheg* on the dialect identification scale, and to *uneducated* on the status scale.
5. In a ZOIB model, the conditional mean of the response is not modeled by a single parameter. Its form is:

$$\alpha_{zoi} \cdot \alpha_{coi} + (1 - \alpha_{zoi}) \cdot \mu$$

where μ is the mean of the beta distribution, α_{zoi} is the zero-one inflation probability, and α_{coi} is the conditional one-inflation probability.

References

- Adank, Patty, Evans, Bronwen G., Stuart-Smith, Jane, & Scott, Sophie K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance* 35(2):520–529.
- Alderton, Roy. (2020). Perceptions of T-glottalling among adolescents in South East England: A sign of 'chaviness,' or a key to 'coolness'? *English Today* 36(3):40–47.
- Anderson, Ragnhild Lie, & Bugge, Edit. (2015). Dialect and other explanatory factors in subconscious verbal guise tests. *Acta Linguistica Hafniensia* 47(2):244–267.
- Auer, Peter. (2018). Dialect change in Europe—Leveling and convergence. In C. Boberg, J. Nerbonne, & D. Watt (Eds.), *The Handbook of Dialectology*. Hoboken: Wiley Blackwell. 159–176.
- Austen, Martha, & Campbell-Kibler, Kathryn. (2022). Real-time speaker evaluation: How useful is it, and what does it measure? *Language* 98(2):e108–e130.
- Avanzi, Mathieu, & Boula de Mareüil, Philippe. (2019). Peut-on identifier perceptivement huit accents régionaux en français? La réponse des sciences participatives. *Glottopol* 31:53–73.
- Beci, Bahri. (1995). *Të folmet veriperëndimore të shqipës dhe sistemi fonetik i të folmës së Shkodrës*. Tiranë: Mihal Duri.
- Beci, Bahri. (2000). *Probleme të politikës gjuhësore dhe të planifikimit gjuhësor në Shqipëri*. Pejë: Dukagjini.
- Bego, Fabio. (2020). Beyond the Albanian–Slav divide: Political cooperation and national identities in the Balkans at the turn of the twentieth century. *East European Politics and Societies* 34(1):25–47.
- Bendixen, Theiss, & Purzycki, Benjamin Grant. (2023). Cognitive and cultural models in psychological science: A tutorial on modeling free-list data as a dependent variable in Bayesian regression. *Psychological Methods* <https://doi.org/10.1037/met0000553>.
- Betancourt, Michael. (2018). Calibrating model-based inferences and decisions. *arXiv* arXiv:1803.08393.
- Bishop, Hywel, Coupland, Nikolas, & Garrett, Peter. (2005). Conceptual accent evaluation: Thirty years of accent prejudice in the UK. *Acta Linguistica Hafniensia* 37(1):131–154.
- Boersma, Paul, & Weenink, David. (2023). *Praat: Doing phonetics by computer*. <http://www.praat.org/>
- Braber, Natalie, Smith, Harriet, Wright, David, Hardy, Alexander, & Robson, Jeremy. (2023). Assessing the specificity and accuracy of accent judgments by lay listeners. *Language and Speech* 66(2):267–290.
- Bugge, Edit. (2018). Attitudes to variation in spoken Faroese. *Journal of Sociolinguistics* 22(3):312–330.
- Bürkner, Paul-Christian. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1):1–28.
- Byron, Janet. (1976). *Selection Among Alternates in Language Standardization: Albanian*. Paris: Mouton.
- Campbell-Kibler, Kathryn. (2011). Intersecting variables and perceived sexual orientation in men. *American Speech* 86(1):52–68.
- Campbell-Kibler, Kathryn. (2012). The Implicit Association Test and sociolinguistic meaning. *Lingua* 122(7):753–763.

- Casini, Laurence, Burle, Boris, & Nguyen, Noël. (2009). Speech perception engages a general timer: Evidence from a divided attention word identification task. *Cognition* 112(2):318–322.
- Cavallaro, Francesco, & Chin, Ng Bee. (2009). Between status and solidarity in Singapore. *World Englishes* 28(2):143–159.
- Çeliku, Mehmet. (1968). Mbi grupimin e të folmeve të gegërishtes jugore dhe mbi disa tipare të tyre. *Studime Filologjike* 5(3):123–136.
- Çeliku, Mehmet. (1971). Kuantiteti i zanoreve të theksuara në të folmet e shqipes. *Studime Filologjike* 8(4):65–100.
- Chen, Hui, Rattanasone, Xu, Cox, Felicity, & Demuth, Katherine. (2017). Effect of early dialectal exposure on adult perception of phonemic vowel length. *Journal of the Acoustical Society of America* 142(3):1707–1716.
- Cheshire, Jenny, & Moser, Lise-Marie. (1994). English as a cultural symbol: The case of advertisements in French-speaking Switzerland. *Journal of Multilingual and Multicultural Development* 15(6):451–469.
- Chiu, Faith, Rakusen, Lindon L., & Mattys, Sven L. (2019). Cognitive load elevates discrimination thresholds of duration, intensity, and f0 for a synthesized vowel. *Journal of the Acoustical Society of America* 146(2):1077–1084.
- Clopper, Cynthia G., & Pisoni, David B. (2004a). Homebodies and army brats: Some effects of early linguistic experience and residential history on dialect categorization. *Language Variation and Change* 16(1):31–48.
- Clopper, Cynthia G., & Pisoni, David B. (2004b). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics* 32(1):111–140.
- Cole, Amanda. (2021). Disambiguating language attitude held towards sociodemographic groups and geographic areas in South East England. *Journal of Linguistic Geography* 9(1):13–27.
- Cole, Jennifer, Steffman, Jeremy, Shattuck-Hufnagel, Stefanie, & Tilsen, Sam. (2023). Hierarchical distinctions in the production and perception of nuclear tones in American English. *Laboratory Phonology* 14(1):1–51.
- Cooper, Robert L. (1975). Introduction: Language attitudes II. *International Journal of the Sociology of Language* 6:5–10.
- de Hoop, Helen, Levshina, Natalia, & Segers, Marianne. (2023). The effect of the use of T or V pronouns in Dutch HR communication. *Journal of Pragmatics* 203:96–109.
- de Vaan, Michiel. (2018). The phonology of Albanian. In J. Klein, B. Joseph, & M. Fritz (Eds.), *Handbook of Comparative and Historical Indo-European Linguistics*. Berlin: de Gruyter. 1732–1748.
- Dickerson, Carly. (2021). *Sociolinguistic Knowledge of Albanian Heritage Speakers in the U.S.* Doctoral dissertation, Ohio State University.
- Dragojevic, Marko, Fasoli, Fabio, Cramer, Jennifer, & Rakić, Tamara. (2021). Toward a century of language attitudes research. *Journal of Language and Social Psychology* 40(1):60–79.
- Dragojevic, Marko, & Giles, Howard. (2016). I don't like you because you're hard to understand: The role of processing fluency in the language attitudes process. *Human Communication Research* 42(3):396–420.
- Dragojevic, Marko, & Goatley-Soan, Sean. (2022). The verbal-guise technique. In R. Kircher, and L. Zipp (Eds.), *Research Methods in Language Attitudes*. Cambridge: Cambridge University Press. 203–218.
- Draxler, Christoph, & Jänsch, Klaus. (2004). SpeechRecorder – A universal platform independent multi-channel audio recording software. In *Proceedings of the 4th International Conference on Language Resources and Evaluation* (Lisbon, Portugal). 559–562.
- Edwards, John. (1999). Refining our understanding of language attitudes. *Journal of Language and Social Psychology* 18(1):101–110.
- Fiske, Susan T., Cuddy, Amy J. C., Glick, Peter, & Xu, Jun. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology* 82(6):878–902.
- Garrett, Peter. (2010). *Attitudes to Language*. Cambridge: Cambridge University Press.
- Garrett, Peter, Coupland, Nikolas, & Williams, Angie. (2003). *Investigating Language Attitudes: Social Meanings of Dialect, Ethnicity and Performance*. Cardiff: University of Wales Press.
- Gjinari, Jorgji. (1968). Diftongjet ua/ue, ie dhe ye në të folmet e gjuhës shqipe. *Studime Filologjike* 5(1):97–106.
- Gjinari, Jorgji, Beci, Bahri, Shkurtaj, Gjovalin, Gosturani, Xheladin, & Dodi, Anastas. (2007). *Atlasi dialektologjik i gjuhës shqipe*. Napoli: Università degli Studi di Napoli.
- Greenwald, Anthony G., McGhee, Debby E., & Schwartz, Jordan L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology* 74:1464–1480.

- Hay, Jennifer, & Drager, Katie. (2010). Stuffed toys and speech perception. *Linguistics* 48(4):865–892.
- Hiramoto, Mie. (2010). Dialect contact and change of the northern Japanese plantation immigrants in Hawai'i. *Journal of Pidgin and Creole Languages* 25(2):229–262.
- Kerswill, Paul, & Williams, Ann. (2002). "Salience" as an explanatory factor in language change: Evidence from dialect levelling in urban England. In M. C. Jones, and E. Esch (Eds.), *Language Change: The Interplay of Internal, External and Extra-Linguistic Factors*. Berlin: de Gruyter. 81–110.
- Kim, Donghyun, Clayards, Meghan, & Kong, Eun Jong. (2020). Individual differences in perceptual adaptation to unfamiliar phonetic categories. *Journal of Phonetics* 81:100984.
- Kingston, John, Baayen, Harald, & Clopper, Cynthia G. (2012). Statistical analyses: Statistics in laboratory phonology mixed-effects models clustering and classification methods. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University Press. 683–691.
- Klatt, Dennis H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59(5):1208–1221.
- Kostallari, Androkli. (1970). Mbi disa veçori strukturore e funksionale të gjuhës letrare shqipe të kohës sonë. *Studime Filologjike* 7(2):3–62.
- Kristiansen, Tore. (2009). The macro-level social meanings of late-modern Danish accents. *Acta Linguistica Hafniensia* 41(1):167–192.
- Kruja, Genti. (2020). Dimension of interfaith harmony in Albanian culture. In *Proceedings of the 19th International RAIS Conference on Social Sciences and Humanities* (New Jersey, USA). 191–199.
- Labov, William. (1963). The social motivation of a sound change. *Word* 19(3):273–309.
- Labov, William. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, William. (2006). *The Social Stratification of English in New York* (2nd ed.). Cambridge: Cambridge University Press.
- Lambert, Wallace E., Hodgson, Richard C., Gardner, Robert C., & Fillenbaum, Samuel. (1960). Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology* 60(1):44–51.
- Lappin, Kerry. (1982). Évaluation de la prononciation du français montréalais. *Revue québécoise de linguistique* 11(2):93–112.
- Lehnert-LeHouillier, Heike. (2010). A cross-linguistic investigation of cues to vowel length perception. *Journal of Phonetics* 38(3):472–482.
- Lenth, Russell. (2023). *Emmeans: Estimated marginal means, aka least-squares means*. <https://CRAN.R-project.org/package=emmeans>
- Liu, Fang, & Eugenio, Evertita C. (2018). A review and comparison of Bayesian and likelihood-based inferences in beta regression and zero-or-one-inflated beta regression. *Statistical Methods in Medical Research* 27(4):1024–1044.
- Llamas, Carmen, & Watt, Dominic. (2014). Scottish, English, British?: Innovations in attitude measurement. *Language and Linguistics Compass* 8(11):610–617.
- Loureiro-Rodríguez, Verónica & Acar, Elif Fidan. (2022). The matched-guise technique. In R. Kircher & L. Zipp (Eds.), *Research Methods in Language Attitudes*. Cambridge: Cambridge University Press. 185–202.
- Matejka, Justin, Glueck, Michael, Grossman, Tovi, & Fitzmaurice, George. (2016). The effect of visual appearance on the performance of continuous sliders and visual analogue scales. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, USA). 5421–5432.
- Milroy, James. (2001). Language ideologies and the consequences of standardization. *Journal of Sociolinguistics* 5(4):530–555.
- Montgomery, Chris, & Moore, Emma. (2018). Evaluating S(c)illy Voices: The effects of salience, stereotypes, and co-present language variables on real-time reactions to regional speech. *Language* 94(3):629–661.
- Moosmüller, Sylvia, & Granser, Theodor. (2006). The spread of standard Albanian: An illustration based on an analysis of vowels. *Language Variation and Change* 18(2):121–140.
- Morgan, Carrie Ann. (2015). *Language Ideologies in Tir'Ona*. Master's thesis, Ohio State University.
- Niedzielski, Nancy. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18(1):62–85.
- Paschen, Ludger, Fuchs, Susanne, & Seifart, Frank. (2022). Final lengthening and vowel length in 25 languages. *Journal of Phonetics* 94:101179.

- Pharao, Nicolai. (in press). Language attitudes and language change. In C. Cieri, K. Drager, L. Hall-Lew, & M. Yaeger-Dror (Eds.), *Dimensions of Linguistic Variation*. Oxford: Oxford University Press.
- Pinget, Anne-France. (2015). *The actuation of sound change*. Doctoral dissertation, Universiteit Utrecht.
- Pipa, Arshi. (1989). *The Politics of Language in Socialist Albania*. New York: Columbia University Press.
- Preston, Dennis R. (1989). *Perceptual Dialectology: Nonlinguists' Views of Areal Linguistics*. Berlin: de Gruyter.
- R Core Team. (2023). *R: A language and environment for statistical computing*. <https://www.R-project.org>
- Riverin-Coutlée, Josiane, Kapia, Enkeleida, Cunha, Conceição, & Harrington, Jonathan. (2022). Vowels in urban and rural Albanian: The case of the Southern Gheg dialect. *Phonetica* 79(5):459–512.
- Roettger, Timo B., Franke, Michael, & Cole, Jennifer. (2021). Positional biases in predictive processing of intonation. *Language, Cognition and Neuroscience* 36(3):342–370.
- Rusakov, Alexander. (2017). Albanian. In M. Kapović (Ed.), *The Indo-European Languages* (2nd ed.). London: Routledge. 552–608.
- Ryan, Ellen Bouchard. (1979). Why do low-prestige language varieties persist? In H. Giles & R. N. St Clair (Eds.), *Language and Social Psychology*. Baltimore: University Park Press. 145–157.
- Ryan, Ellen Bouchard. (1983). Social psychological mechanisms underlying native speaker evaluations of non-native speech. *Studies in Second Language Acquisition* 5(2):148–159.
- Scharinger, Mathias, Monahan, Philip J., & Idsardi, William J. (2011). You had me at “Hello”: Rapid extraction of dialect information from spoken words. *NeuroImage* 56(4):2329–2338.
- Sims, Ronald R. (2002). *Managing Organizational Behavior*. London: Quorum Books.
- Stoet, Gijsbert. (2010). PsyToolkit—A software package for programming psychological experiments using Linux. *Behavior Research Methods* 42(4):1096–1104.
- Stoet, Gijsbert. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology* 44(1):24–31.
- Trudgill, Peter. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society* 1(2):179–195.
- UNDP. (2018). *Religious tolerance in Albania*. <https://www.undp.org/sites/g/files/zskgke326/files/migration/al/English-report-religious.pdf>
- Vaicekauskienė, Loretta. (2019). Driving forces behind language change. Does Danish theory hold up in Lithuania? *Scandinavistica Vilnensis* 14:241–272.
- van Osch, Sylvie M. C., & Stiggelbout, Anne M. (2005). Understanding VAS evaluations: Qualitative data on the cognitive process. *Quality of Life Research* 14(10):2171–2175.
- Watson, Kevin, & Clark, Lynn. (2015). Exploring listeners' real-time reactions to regional accents. *Language Awareness* 24(1):38–59.
- Wentzky, Chase, & Summers, Joshua D. (2020). Individual differences in describing levels of automation. In *International Design Engineering Technical Conferences & Computers and Information in Engineering Conference* (St. Louis, USA). 19136.
- Winter, Bodo, Fischer, Martin H., Scheepers, Christoph, & Myachykov, Andriy. (2023). More is better: English language statistics are biased toward addition. *Cognitive Science* 47(4):e13254.
- Xhudo, Gus. (1995). Tension among neighbors: Greek-Albanian relations and their impact on regional security and stability. *Studies in Conflict & Terrorism* 18(2):111–143.
- Zahn, Christopher J., & Hopper, Robert. (1985). Measuring language attitudes: The speech evaluation instrument. *Journal of Language and Social Psychology* 4(2):113–123.

Cite this article: Riverin-Coutlée, Josiane, Enkeleida Kapia and Michele Gubian. (2024). Dialect change and language attitudes in Albania. *Language Variation and Change* 1–24. <https://doi.org/10.1017/S0954394524000103>