Balti Dialectology and Variation

Jennifer Kaplan

1. Introduction

The following study of Balti dialects has two parts: First, to compare the dialect of a contemporary Balti speaker to prior dialectological surveys in order to judge both their thoroughness and any potential linguistic shifts; and second, to collect sociolinguistic data on attitudes toward the Balti language and Balti speakers.

The first, comparative portion of this project uses data from Peter Backstrom's (1992) survey of Balti, with data collected in 1989-1990. Peter Backstrom recorded 205 words from speakers across seven regions of Baltistan: Rondu Valley (RON), Shigar Valley (SHG), Skardu Valley (SKB), Khapalu Valley (KHP), Kharmang Valley (KMG), Chorbat (CHO), and Skardu Purki (PSK). All are dialects of Balti except for Skardu Purki, which is a dialect of the closely related language Purki.

Backstrom recorded that 50 of these words, or nearly 25% of the basic vocabulary, was exactly the same in all seven regions of Baltistan—meaning that there were no lexical or phonological differences for 50 of the words he collected. The remaining data reveal dialectical variation on the level of phonetic production and, to a lesser extent, lexical differences. He proposes that all six of the Balti regions surveyed have a very high degree of lexical similarity, averaging around 90% (Backstrom 1992: 11). There were more significant differences between the Balti dialects and Purki, though mutual intelligibility was still high. Further, there is far more phonological variation between the various dialects of Balti than there is lexical variation.

The purpose of the *comparative* portion of this study is to match the word forms our consultant uses against prior recorded forms of Balti, using Backstrom's data as a basis. More specifically, this study compares a modern-day Balti speaker's language to previously recorded forms of Balti dialect variation as collected by Backstrom (1992). This study seeks to critically examine 1. The accuracy of the prior dialectical data and, more broadly, 2. how a cosmopolitan speaker's speech might blend various dialectical features.

This paper begins with an overview of the methods used to collect dialectical and ethnographic information, as well as an explanation of dialect-matching. Section 3 goes over the data in more detail with a breakdown of how the data in this study matches up against the data collected by Backstrom (1992). This section includes subsections on lexical variants not documented in Backstrom's study, connections between the consultant's personal history and the dialectical variants he uses most often, and a summary of the issues that resulted from relying on Backstrom's data for a description of Balti dialect continua. The paper concludes with recommendations for future research.[1]

[1] As a formatting note, numbers in parentheses refer to the numbering in Backstrom (1992)'s word list.

2. Methods

This paper has two tasks: Comparing our consultant's speech to prior records of Balti dialects, and collecting data on linguistic attitudes that Balti speakers hold towards different dialects of their language.

For dialectical analysis, this paper relies primarily on the dialect continua in Backstrom (1992) to match our consultant's speech forms to dialects throughout Baltistan. The choice to rely on Backstrom's paper for a description of various Balti dialects was primarily a result of necessity, as very little work exists on dialectical variation in Balti.

The primary challenge in calculating how our consultant's dialect compares to previously documented regional dialects of Balti involved deriving a system for matching our data to previously collected data. The system I came up with assigned one of three degrees of similarity to each word (token) in the 205-strong wordlist listed by Backstrom (1992): *Exact Match, Close Match,* or *No Match.* A complete list of every data point collected is included in **Appendix 3,** where *Exact Matches* are blue, *Close Matches* are purple*,* and *No Matches* are red.

After determining how or whether a token or tokens used by our consultant matches one or more of the forms in Backstrom's (1992) list, I assigned point values to each vocabulary term based on degree of match. Each word has a point value of '1.' A word that was either an exact match or a close match with one of the word forms Backstrom elicited is then added as one 'point' for each region where that specific form was elicited. For example, (48), 'cloud' is recorded as [psɪl] in RON, SHG, SKB, and KHP, as [ɸsil] in KMG and CHO, and [sɪl] in PSK. Because our consultant uses the form [ɸsil], the regions KMG and CHO were each awarded one point for this word. After all words had been compared to Backstrom's data, the total scores for each dialect were added up.

In order to collect data on linguistic attitudes among Balti speakers, I conducted a series of ethnographic interviews with our consultant, Muhammad. In the first interview, included here in **Appendix 1,** I focused on Muhammad's family history in Baltistan, and his time spent living in various regions. The second interview, included here as **Appendix 2**, switched gears to focus more narrowly on linguistic stereotypes of speakers of different Balti dialects.

3. Data & Analysis

The first part of this deals primarily with the word list data collected from our consultant, modelled on the word list that Backstrom used. A complete list of the data can be viewed in **Appendix 3.**

I begin by overviewing the data, before analyzing word forms our consultant uses that were unrecorded in Backstrom's data. I then move into a sociolinguistic analysis of our consultant's speech using information from my ethnographic interviews, focusing on language contact and the various regional influences potentially at play in our consultant's speech. The final

subsection under this heading outlines problems with Backstrom's data that created problems for my comparative analysis.

3.1 The Data at a Glance

Of the 205 words in Backstrom's list, we have data for 200. Here is a quick rundown:

> Number of Words that Matched Exactly with Backstrom's Phonological Transcriptions: 101/200[2]
> Number of Words that Matched Closely with Backstrom's Phonological Transcriptions 86/200
> Number of Words that did not match Backstrom (to be elaborated on): 13/200

Among the words that matched closely, but not exactly, with Backstrom, the most frequent point of contention was in the transcription of vowels. This was not unexpected, as the nature of vowel production makes their transcription generally more open-ended than consonant transcription (Ladefoged & Johnson 2010).

In short, one of my hypotheses going into this project was that Muhammad's speech would contain features of multiple Balti dialects, precisely because he not only has travelled to every reason of Baltistan, but also because he has lived for extended periods of time in multiple regions—including cosmopoilation regions (such as Skardu Valley) where he had prolonged contact and developed relationships with individuals from across Baltistan. Given the data, this hypothesis has proven correct. As Section 3.2 illustrates, a plurality of Muhammad's forms match those from either Kharmang or Skardu, those his vocabulary includes forms that are unique to each of the seven geographical regions that Backstrom surveyed.

3.1.1 Matchless Words

As previously mentioned, there were 13 words in Backstrom's set that had no dialectical match with Muhammad's productions; some of these involved significant phonological departures from Backstrom's transcriptions, and are entirely separate lexical items . I have separated these thirteen words into three groups: Firstly, Those that are close enough to indicate either A. A possible phonological shift in Balti, and B. Transcription error. Secondly, words whose differences from Backstrom's data set indicate potential linguistic broadening or narrowing, and 3. Words for which I can offer no explanation for their distinct departure from any form recorded by Backstrom.

TABLE 3.1.1.1: 'Close, but no cigar'

	Consultant Production	**Backstrom's data**
Bone	[rulspɛ]	[ruspʌ]; [ruspa]

[2] Note that Backstrom's *numbering* goes up to 210, but since he omits several words from his final list, the total number of words is *less* than 210

Roof	[tokʰ]	[hʌndoq]; [handoq]; [tʰoqsa]
Woman	[bʊrsiŋ]	[bustriŋ], [brustiŋ], [bʊstriŋ]
Twelve	[tʃuŋəs]	[tˆʃoŋɛs], [tˆʃoŋʌs]
Cow	[bzo:]; [bu:]; [bo:]; [bzo:mo]; [ban]; [ɬaŋ]	[bʌ], [bʌlʌŋ], [ba].
Elbow	[tˆʂɪmohs]	[tˆʂɪmok], [tʰrɪmoŋ], [tʰrɪmok], [pʊrtˆsʊk], [kʰriõŋ], [kʰrimoks]

In TABLE 3.1.1.1 we see a list of words that differ from Backstrom's data, but that are similar enough to the 1992 data that they are likely still related to Backstrom's forms. Some, such as *bone* or *elbow*, may differ due to transcription errors. Others, such as *woman*, seem to be undergoing metathesis that is observable elsewhere in the data. We also see words like *roof* which appear to be shortened in Muhammad's dialect. Further, there is *twelve,* which is quite close to Backstrom's data but differs in the final vowel and, thus, choosing one of the two variants that he records would be arbitrary. Issues with data collection for the words for *cow* are discussed in Section 3.4.

TABLE 3.1.1.2: 'Potential Linguistic Broadening'

	Consultant Production	**Backstrom's data**
Child	[balbis]	[pɭu], [pʰru]
Sister, younger	[aʃʒo]	[striŋmo], [noŋo], [nõŋo]

Moving on, in TABLE 3.1.1.2 we see words recorded with more than one meaning in the dictionary. The word Muhammad offered for *child*, [balbis], is also used for 'baby'; similarly, his word for *younger sister*, [aʃʒo], is also used as a general term for 'sister.'

TABLE 3.1.1.3: 'Words with no Clear Connection to Prior Data'

	Consultant Production	**Backstrom's data**
Village	[dɾul]	[jul], [gʷroŋ], [groŋ], [ʒoŋ], [dˆʒoŋ],
Thread	[tˆsɛ.ma]	[skutpa]; [skutbʌ]; [skutpʌ]
Cloth	[gʊŋtˆʃəs]	[kʌpɾʌ], [kʌpɾa], [snambu], [rʌs]
Mud	[ldʌm]	[dˆʒ^j agʌt], [dˆʒagʌt], [tˆʃʰɪn,] [dˆʒaldˆʒɪt], [ldˆʒakldˆʒɪt], [lʌmtˆʃʌr], [ldok]
Short	[tsʊntsɛ];[titə]	[tˆʃʌt], [tˆʃʌttˆʃʌt]

Finally, TABLE 3.1.1.3 contains words that do not seem to be related to any of the forms collected by Backstrom. Nor are these merely lexical alternatives. When asked about the word for 'short,' for example, Muhammad indicated that he had never heard the forms [t͡ʃʌt] or [t͡ʃʌtt͡ʃʌt]—the only two versions Backstrom recorded across any of the seven regions of Baltistan. Such inconsistencies indicate one of three scenarios: 1) The consultant has been outside of Baltistan long enough that terms he would have used more rarely has fallen out of his vocabulary; 2) The forms Backstrom recorded for "short" actually mean something else; 3) The terms Backstrom recorded in 1992 are no longer in use nearly 25 or so years on, either due to lexical shift or for other reasons.

3.2 Cosmopolitan Language Contact in Our Consultant's Speech

Muhammad has lived in multiple regions of Baltistan (See **Appendix I** for a full account of where he has lived and his metalinguistic observations about the dialects of Baltistan). After having been born and raised in the region of Kharmang, he went to the city of Skardu from the ages of 16-18 to complete his secondary education. In Skardu, he came into contact with individuals from a variety of ethnic and linguistic backgrounds; according to him, people from every region of Baltistan come to live in Skardu. According to theories of sociolinguistics, 17 is the age when we are most linguistically impressionable and, consequently, the most linguistically innovative (Tagliamonte & D'Arcy 2009); considering that Muhammad spent his teenage years in a cosmopolitan city and spent time among a regionally linguistic diverse group of friends, it would make sense that his vocabulary will not be limited to the patterns of only one region, but rather represent a fusion of the various linguistic groups he was in contact with.

3.2.1 Between Skardu and Kharmang

Unsurprisingly, the majority of the forms used by Muhammad are recorded in either Kharmang or Skardu. More specifically, 157 of the 205 words elicited matched Backstrom's recordings for forms used in either Skardu or Kharmang. This makes sense, due to Muhammad's extensive time living in both regions (See Section 3.2).

This means that 48 terms match *neither* Skardu or Kharmang. Of these, 18 of the forms Muhammad uses do not match any of the dialects recorded by Backstrom; the remaining 30 matched a mix of forms from Rondo Valley, Shigar Valley, Khapalu Valley, Chorbat, or Skardu Purki. A number of phonological forms that Muhammad uses are recorded by Backstrom as only occurring in one specific dialect; for a list of the number of forms unique to a given region that Muhammad uses, see TABLE 3.2.1.

TABLE 3.2.1: 'No. of Phonological Forms Unique to a Given Region Appearing in Consultant's Speech'

Rondo Valley (RON)	1
Shigar Valley (SHG)	4
Skardu Valley (SKB)	4
Khapalu Valley (KHP)	6

Kharmang Valley (KMG)	4
Chorbat (CHO)	3
Skardu Purki (PSK)	4

It is not possible to say with any certainty *why* Muhammad uses some phonological forms that Backstrom records as only occuring in regions where he has never lived, though we can theorize a few different explanations. Firstly, it's possible, as I have been suggesting in this section, that some of Muhammad's phonological realizations were influenced by his contact with speakers from across Baltistan while living in the cosmopolitan city of Skardu. Secondly, it's possible that the phonological forms Backstrom recorded were not quite as regionally distinct as he hypothesized; Backstrom primarily relied on a single speaker as a representative of each region, and we know from Muhammad that Balti speakers from different regions do have quite a bit of contact with one another, depending on their regions of origin (some regions are more isolated than others due to geographical barriers).

3.3 Potential Lexical Shift

Some of the discrepancies between Backstrom's data and the data from this project hint at potential lexical shifts ongoing in Baltistan.

At the nominal level, there are two notable examples. Backstrom lists the form [trobar] under 'noon' (120); however, in Muhammad's dialect, [trobar] means *lunchtime*, with the word for 'noon' being [pʰɪʃiŋ]. This is possibly an instance of lexical *narrowing* in Muhammad's dialect. On a slightly different level, Backstrom records the word [paju] as meaning 'salt' in all Balti dialects, which corresponds with Muhammad's use of the word. However, Backstrom records the word for 'salt' as being [t͡ʃʰa] in Skardu Purki. This appears to be a misinterpretation of the data, as [t͡ʃʰa] means *tea* in Muhammad's dialect in Balti—and, indeed, the word is a cognate across a wide swath of Sino-Tibetic languages and has been borrowed into English as *chai*. The confusion here seems to be in the commonality of a specific type of beverage, *salted tea,* [paju t͡ʃʰa]—Indeed, when first asked about the word for *salt* on its own, Muhammad offered an anecdote about how common it is to put salt in one's tea. While it is possible that the common association of these two food items—*salt* and *tea*—has led to the word for 'tea' standing in for the word for 'salt,' the importance of both items in Baltistan renders this kind of phrasal metathesis, or *synechdoche*, unlikely. More plausible is that Backstrom's Skardu Purki consultant mentioned [t͡ʃʰa] as something they associated with *salt*.

More interestingly, our data hints at a potential shift in the use of some *function words* in Balti. Backstrom lists the word [boŋo] as a variant of the pronoun 'she' (206) in Skardu Purki. In Muhammad's dialect, [boŋo] means 'daughter.' Backstrom also records [boŋo] as meaning 'daughter' (112) in all seven of the varieties he covers. It's *possible* that [boŋo] has been reevaluated as a pronoun in PSK. This pattern of lexical shift is not unheard of; the French pronoun *on* comes from Latin *homme* (on < om < homme) (Richard 1989: 23). However, given Backstrom's own hesitance in deeming his Skardu Purki consultant reliable (he gave the

consultant a ‘C’ grade), we’d need more data before we can confirm that a shift in meaning is occurring in this variant.

3.4 Issues with Backstrom’s Data

There are several issues with Backstrom’s data that makes it less than ideal as a point of comparison with our consultant’s speech. A reliance on *word lists* limits both the reliability of the forms Backstrom collects, while his analysis could have gone further in explaining why he chose to list multiple versions of a particular word in some cases, and only one version in other cases. Backstrom himself acknowledges that his data may not be super accurate, and thus devises a rating system (A-C) for his consultants; the unreliability of some data becomes more apparent upon closer analysis.

3.4.1 Solely Relying on Word Lists

Firstly, Backstrom collected most of his data using word lists; such collection provides far shallower information about a given language than writing a sketch grammar or compiling a dictionary. The limits of the word-list method are most apparent in the forms that he collects, wherein word forms recorded for some dialect regions are missing case- or tense-markings, but others are not; for example, in listing the various dialectical forms of “he went,” forms for SKB, SHG, KHP and CHO are listed with the past marker *-s*, while forms for the other regions do not have the past-marker. We know from working with Muhammad that the past marker is often dropped in unguarded speech—this is likely the phenomenon that Backstrom covered. However, it’s unclear, exactly, how these forms were collected; Muhammad uses both [soŋs] and [soŋ] for *he went* depending on the formality of his speech, and it’s likely that other consultants would also vary in their dropping of the tense marker. Thus, it would make more sense to specify both 1. The significance of consultants in some regions dropping the *-s* marking, and 2. Whether the dropping of tense-markers was a consistent feature in consultants’ speech, or if they varied in dropping versus retaining tense-marking in different linguistic contexts. Backstrom’s data would have likely been clearer had he elicited vocabulary in different contexts, rather than solely relying on word lists.

3.4.2 Lack of Specificity in Word List

Additionally, Backstrom doesn’t clarify more specific meanings for the words on his list. For example, the forms he lists for (45), ‘rain,’ are [t͡ʃʰʌrpa], [t͡ʃʰʌrɸʌ], [t͡ʃʰʌrfʌ], [t͡ʃʰʌrpʌ], [t͡ʃʰʌrfa], and [t͡ʃʰarɸa], which are all clearly variants of the same underlying word.

However, during our earliest elicitation sessions in January 2020, when asked for the word for *rain* in isolation, Muhammad offered [namkʊɾ], which he later clarified meant ‘soft rain.’ Over the subsequent months, Muhammad offered both more specific and broader words for *rain*: One mighy say that it was ‘cloudy,’ [kʊɾɛ], that there was a ‘storm,’ use a verbal form to clarify that it has ‘started raining’ [ɾjasɛt] or ‘started lightning’ [ɬok baɾba ɾjasɛt], or, most generally, simply refer to the ‘weather’ [nambzo]. In fact, it was only when I began working specifically

from Backstrom (1992)'s data and asked our consultant for a word that started with the affricate [t͡ʃʰ] that he offered [t͡ʃʰarɸa], which he specified as meaning 'aggressive rain,' rather than rain in a general sense. One may be tempted to claim to claim that Baltistan's mountainous climate leads to multitudinous words for similar phenomena (a la the oft-cited example of Inuit having 'fifty words for snow'), and thus Backstrom could not have predicted the need for greater specificity in drafting his vocabulary list; however, it should be noted that English, too, has nuanced linguistic gradations for describing weather (note a *drizzle* versus a *downpour*) and thus this is no excuse. The lack of clarification regarding which words were targeted makes it difficult to gage the representativeness of the data collected.

A similar process happened with (89), 'cow.' For this word, Backstrom lists data for only four of the seven regions: his SHG consultant uses [bʌ], as does his SKB consultant in variation with [bʌlʌŋ], the same form produced by his Kharmang consultant, while his CHO consultant uses [ba]. The rows for RON, KHP, and PSK are all marked by three dashes, indicating that he found no suitable data for those regions. 'Cow' is one of only four words in Backstrom's entire 200-plus word set to have data missing from any region, and the only word with data missing from three regions. It's not possible that Balti speakers from these three regions have no word for 'cow'; livestock, particularly cattle, are an integral part of Baltistan's agricultural economy (and so common in the region that people often have problems with cows breaking down the doors of their homes, according to Muhammad). Rather, I propose that the problem lies in Backstrom's methodology: Trying to find a singular word for *cow*. Muhammad uses [nʊɾban] to describe livestock generally, which he claims is a portmanteau of [nʊɾ], 'sheep' and [ban], 'cow.' One can also use [ɬaŋ] or [bzo] for 'male cow,' [bzo:mo] for 'female cow' and [bu:] for 'calf,' while the term [ɬuŋse] which specifically means 'livestock lying down.'

If there *is* an interesting linguistic choice behind the elicitation of this specific word for rain (e.g., it's taken on a *broader* meaning in parts of Baltistan, encompassing rain in a general sense), this is never mentioned in Backstrom's analysis.

4. Linguistic Attitudes in Baltistan

My interviews with Muhammad also revealed some of the linguistic attitudes that Balti speakers have towards their language and various Balti dialects. I begin with a description of the association between language and identity, before moving into a description of multilingual influences in Baltistan and ending with a description of linguistic stereotypes.

4.1 Ethnic and Linguistic Correlates

Just as Backstrom found, I found in my interviews with Muhammad that *Balti* is used interchangeably as a linguistic and an ethnic identifier—In fact, linguistic identity heavily correlates with ethnic identity throughout Muhammad's descriptions of the population of Baltistan. Despite the high prevalence of bilingualism in Muhammad's descriptions, those who are ethnically Balti primarily self-identify as Balti speakers. Muhammad himself fits this pattern; despite being multilingual, his primary identity is that of a "Balti speaker."

4.2 Interlingual Mixing

The prevalence of multilingualism, especially in the cities of Baltistan, has led to a large amount of linguistic interchange. Our consultant is aware of many borrowed words in his speech, and has highlighted words of Urdu, Arabic, Hindi, Persian, and English origins in common use among Balti peoples. According to Muhammad, English has taken on increasing importance in the education system as a language of globalization; many of the words for common school items, such as 'pencil' and 'table,' are borrowed from English.

As Backstrom notes, there is also a large amount of mutual comprehension between Balti speakers and Purki speakers, as well as Balti speakers and Ladakhi speakers.

4.3 Comprehension Differences Across Dialects

Consistent with Backstrom's findings, Muhammad indicates that there is a high degree of mutual intelligibility in the Balti across Baltistan. When asked if there were any dialects he had difficulty understanding, he mentioned only the Balti in Ganche, mostly due to its distinctive vocabulary and slang.

4.4 Perceptions of Dialectical Purity

In a phenomenon related to the conflation of ethnic identity with linguistic identity, the "purity" of a given region's Balti is judged based on its proportion of Balti (wherein *speakers* and members of the ethnic group are interchangeable). Though according to Muhammad every region thinks that their Balti is the most "correct," this population/purity correlate seems to predominate above individual regional pride. He considers both Shigar and Ganche to be "100% Balti" regions; this contrasts from Backstrom's findings that Khapalu was most frequently cited by native speakers for its purity (Backstrom 1992: 22). It's not just inhabitants that can impact the perceived "purity" of a given region's Balti; Muhammad also mentioned that the large influex of both national and international tourists into Shigar due to K2 being located in the region may result in inhabitants' Balti being less pure, perhaps due to the influence of tourists' dialects.

On the opposite end of the spectrum, our consultant associated a lower degree of Balti purity with multi linguistic regions. He described the Balti from his own region of Kharmang as mixed with the Shina language; as with Shigar, he also implied that the region's recent opening to tourism has also decreased the purity of the Balti spoken there. Like Kharmang, he considers the Balti of Skardu to be less "pure" and more mixed, though still thinks that the region is about "80-90% Balti." As mentioned previously, Skardu houses the capital of Baltistan and is also the most cosmopolitan region, with many Shina and Urdu speakers.

Throughout Muhammad's answers about linguistic purity, I noticed a subtle but important association between two different types of "purity." A region's Balti may be considered more pure because 1. There are no other languages spoken in the region, and thus less linguistic borrowing, or 2. The region is solely populated by ethnic Balti, who speak the Balti region. Once

again, we see that the conflation between ethnic and linguistic identity is important to Balti speakers' language ideologies.

4.5 Dialectical Stereotypes

I asked Muhammad directly if he was aware of any stereotypes about the Balti spoken by people from different regions. He spoke of two different types of stereotypes: First, a more general distinction between speakers from different geographical regions (e.g., "river" people and "mountain" people), and second, of more specific linguistic stereotypes regarding people from the different regions of Baltistan.

Baltistan is a mountainous region filled with valleys and rivers, bordered by the Himalayas. The Indus River runs between the mountain; 98% of Balti speakers live by the river. Some people also live up in the mountains [ʃina]; these people are called [bɻok-pʰa], with [bɻok] etymologically related (according to Muhammad's intuition) to the word for mountain, [ɻi], and the whole first part of the phrase loosely translating to a sense that "people live in___", with [pʰa] translating to "nationality." Muhammad considers these people "more rural" than those who live by the river.

Those who live by the river feel that they are more "civilized," though this perception has not led to any "racial" divisions between the groups; rather, it has resulted in the spread of certain "stereotypes" about each group (mountain versus river Balti speakers). People who live by the riverside feel more "strong," "powerful," and "emotional." Among those who live in the mountain, there is the perception that the river people "make a lot of mistakes" and are more "arrogant." When asked about any perceived differences in the accents of the river people versus the mountain people, Muhammad commented that the mountain people will use [g] where river people would use [ɻ], leading to such forms as [ga-ʃiŋ] versus Muhammad's preferred form, [ɻa-ʃiŋ].[3]

When asked about any stereotypes about Balti speakers in Kharmang in particular, Muhammad discussed their supposed propensity for speaking quickly, being active, and lengthening their vowels. Because those from the Kharmang region often have to travel to other parts of Baltistan for work opportunities, people from the Shigar Valley refer to Kharmang people as [bjoŋbu], 'the fly'—speaking quickly, travelling often, and arriving everywhere. Moreover, those in Shigar will mock Kharmang Balti speakers' vocalic pronunciations, exaggerating their tendencies to lengthen vowels.

When describing stereotypes about speakers from Shigar, Muhammad indicated that they are perceived as having a different prosodic pacing with their pronunciations, and lengthen word-final or sentence-final sounds in comparison to the rest of their speech.

In describing Skardu speakers, Muhammad referenced the cosmopolitan nature of the city. Different parts of the city have different pronunciations. Neighborhoods are differentiated by ethnic backgrounds, but less distinct than New York City—a frequent point of comparison in our

[3] I do not have a translation for this word, but it looks like it might be 'goat meat.'

discussions, as Muhammad now lives in Queens. The most important question in Skardu is ‘From what village do you come?, with an overall more microcentric sense of belonging. This description of Skardu’s micro communities corresponds to common sub-communal linguistic ideologies that occur in many urban environments; for example, the important conception among New Yorkers that Staten Islanders, Brooklynites, Manhattanites, Bronxites, and those from Queens all have distinctive accents is important to linguistic identity-building, even though phonological surveys have confirmed that there are no distinct accents that differentiate speech from the boroughs (White 2020). Thus, we can consider the conception that Balti speakers in different neighborhoods of Skardu have different “accents” to be an extension of linguistic community and identity-building.

5. Limitations of this Study and Further Directions

There are several limitations that a future study could address. Firstly, a better comparative method would be able to account for *degree of closeness* in cases where a given phonological form doesn’t quite match one of Backstrom’s. For example, cases I would consider ‘minor difference’ include a single difference in vowel transcription, such as (101) [mintax] versus Backstrom’s [mintʌx] for ‘name.’ Coupled with this limitation, it would be useful to have a way of accounting for how close *other dialects* are to the form I ultimately chose as ‘matching’ our transcription. In some cases, other dialects of Balti appear to have very minor differences from one another (such as [stare] versus [stʌre], ‘axe’), whereas in other cases, the differences can be quite drastic (Muhammad’s use of [bɛbʒʊŋ] for ‘egg’ versus Backstrom’s closest dialectical transcription being [biʌpʒʊn]). To ultimately judge which dialect our consultant’s speech most closely matches (or, more accurately, which dialects he is blending), it would be necessary to create a separate proximity scale for the differences *between* dialects.

References

Backstrom, P. C. & Radloff, C. F. (1992). Sociolinguistic Survey of Northern Pakistan, Volume 2: Languages of Northern Areas. National Institute of Pakistan Studies and Summer Institute of Linguistics: Islamabad.

Ladefoged, P. & Johnson, K. (2010). *A Course in Phonetics* (6th ed.). Cengage Learning.

Payne, T. E. (1997). *Describing Morphosyntax: A Field Guide for Linguists.* Cambridge, UK: Cambridge University Press.
Rickard, P. (1989). *A History of the French Language.* Taylor & Francis Group.

Tagliamonte, Sali A. and D'Arcy, A. (2009). "Peaks Beyond Phonology: Adolescence, Incrementation, and Language Change." *Language* 85, (1) 58-108.

White, E.J. (2020). *You Talkin' to Me? The Unruly History of New York City English.* Oxford University Press.

Appendix I: Ethnographic Interview with our Consultant, Muhammad Hussein

All information is taken from a meeting with Muhammad on July 30th, 2020. The first two questions are borrowed from Payne (1997). Questions and answers have been edited for clarity.

Broad Language Background: Questions from Payne 1997

1. Q: What term do the people use to distinguish themselves from other language groups?

Baltistan is a mountainous region filled with valleys and rivers, bordered by the Himalayas. The Indus River runs between the mountain; 98% of Balti speakers live by the river. Some people also live up in the mountains [ʃina]; these people are called [bɻok-pʰa], with [bɻok] etymologically related (according to Muhammad's intuition) to the word for mountain, [ɻi], and the whole first part of the phrase loosely translating to a sense that "people live in___", with [pʰa] translating to "nationality." Muhammad considers these people "more rural" than those who live by the river.

2. Q: What is the nature of the interaction with these language groups? Economic? Social? Friendly? Belligerent?

Those who live by the river feel that they are more "civilized," though this perception has not led to any "racial" divisions between the groups; rather, it has resulted in the spread of certain "stereotypes" about each group (mountain versus river Balti speakers). People who live by the riverside feel more "strong," "powerful," and "emotional." Among those who live in the mountain, there is the perception that the river people "make a lot of mistakes" and are more "arrogant." When asked about any perceived differences in the accents of the river people versus the mountain people, Muhammad commented that the mountain people will use [g] where river people would use [ɻ], leading to such forms as [ga-ʃiŋ] versus Muhammad's preferred form, [ɻa-ʃiŋ].[4]

Muhammad's Background

1. What part of Baltistan are your parents from? Your grandparents?

Baltistan is subdivided into provinces; it lies in an area that is part of the Kashmir dispute. There are ten districts in gilgit-Baltistan (the larger administrative area), and four districts in Baltistan: Skardu, Kharmang, Ganche, and Shigar. Muhammad is from Kharmang, which he says translates to "many forests." His family has lived in Kharmang for at least four generations.

2. Where in Baltistan were you born (i.e., region)?

Muhammad was born in Kharmang (see above).

3. Before college, where did you go to school?

[4] I do not have a translation for this word, but it looks like it might be 'goat meat.'

Muhammad went to school about ten km from where he grew up.

4. Where did you attend university? In what region is this university located in?

From the age of 16, he attended an "inter-college" in Skardu. For the next two years, he went to the central Madrasa for Shia muslims—the University of the Punjab—located in Lahore, Punjab, Pakistan. He spent a total of four years here studying both religious and secular education.

For the next part of his education, he spent four years in post-grad studies in Islamabad, receiving degrees in both Islamic Studies and Islamic History and an MA in Educational Planning.

5. In which parts of Baltistan have you lived (any/every village, region, etc.)? How long did you live in each place?

He has lived in Kharmang and Skardu.

6. Where were your school friends from? What languages did they speak at home?

When he grew up in Kharmang, all of his friends were from Kharmang. However, Skardu is much more urban than Kharmang; people living here come from all four regions of Baltistan. Thus, Muhammad's friends from his inter-college years in Skardu came from across Baltistan. Further, when he began pursuing higher education abroad, his friends were similarly from diverse national and linguistic backgrounds.

7. What languages did you speak with your school friends?

Muhammad spoke Balti with his school friends.

8. How much have you travelled in Baltistan?

By his own account, he has visited most parts of Baltistan. Most of his time in Skardu province was spent in Skardu city, and he was in Ganche briefly for educational reasons.

9. What region (or village) in Baltistan feels most like "home"?

Kharmang is first and foremost the place that feels most like home for Muhammad, with Skardu coming in second place.

    a. Do you feel like you sound like other people from this place?

Both Kharmang and Skardu have mixed populations, and are not monolingually Balti-speaking. By contrast, he considers both Shigar and Ganche to be "100% Balti" regions. He identifies as Balti/a Balti speaker.

Appendix II: Second Ethnographic Interview with Our Consultant, Muhammad
All information is taken from a meeting with Muhammad on August 18th, 2020. Questions and answers have been edited for clarity.

1. What's the stereotype of what people in Khamang sound like?

[They're like] [bjoŋbu], 'the fly.' These people arrive anywhere, travel a lot, are socially active. Kharmang is more narrow than other valleys in Baltistan, has fewer resources, so people travel to other areas for jobs and work. Shigar Valley is wider, they call Kharmang people [bjoŋbu].

Shigar people imitate Kharmang people by making words and whole sentences sound longer, especially mocking (words like) [t͡ʃi:::bɛs], with different pacing in the syllables

2. What are some other regional stereotypes?

In Shigar, (they say) [le kaŋ on-sa::]—longer at the end of, or on one end of, a sentence, where the rest of the sentence is said faster.

In Kharmang, everything is faster.

In Skardu, they are called [ʒilos], 'jealous,' [baŋjɛt] 'jealous with a negative intention,' or [bɛtnjɛt], a similar word borrowed from Arabic or Perisan; because migrants are wealthier in Kharmang, people in Skardu sold their lands. They have a specific way of speaking, but it's hard to describe! Different parts of the city have different pronunciations. Neighborhoods are differentiated by ethnic backgrounds, but less distinct than New York City. [The more important question is] 'From what village do you come?' There's a more microcentric sense of belonging.

3. Which region has the "purest," or "most correct," Balti? Which region has the "least pure" or "least correct" Balti?

Everyone says they do! Maybe Ganche?

In Kharmang the Balti is mixed with Shina language; tourists were not allowed for many years. Now it's opened up to tourists, mostly from Pakistan, many from Europe and the Middle East.

Shigar and Ganche are purely Balti. There is no other ethnic of language group [in these regions], but Shigar has a lot of tourists because of K2.

Skardu has all regional languages, but is 80-90% Balti; [they also speak] Shina, Urdu, [and] the village of [stəɾmʌns] is more Urdu.

4. Are there any regions or groups whose Balti is harder for you to understand?

It can be difficult to understand the Balti in Ganche, which has distinct vocab and slang. It sounds a little different, mostly the vocab.

5. How far is your village from Skardu? What are the distances between the various Balti regions?

Muhammad's village is 60 km to Skardu; from the other sode of Kharmang, it's 200 km. To the border of Gilgit, it's 200 km. To Shigar/Ganche, it's about 1000 km.

6. How do you travel between regions?

You have to fly from Islamabad to Skardu. Everything else is connected by road networks, boats, [gulaɾi], 'chairlift'.