

Progress Report No. 1: Phonology of Balti

Jessica Kuleshov, Columbia University

1. Introduction

Balti is a language spoken in Pakistan, and is a language with about 290,000 speakers. It is spoken in the Balti part of the Gilgit-Baltistan region, which is part of the Kashmiri conflict, and the language and the culture have no formal recognition by the Pakistani government. It belongs to the Tibeto-Burman language family. The elicitation of the phonemic inventory and phonological rules and all reports henceforth are based on elicitations from Muhammad Hussein, a native speaker of Balti currently residing in the United States, and as such all results are based off of his idiolect.

In this paper I will discuss the phonology of Balti. §2 will discuss the consonants, §3 discusses vowels, §4 discusses allophones, §5 discusses phonotactic constraints, §6 discusses phonological rules that I have been able to so far determine, and §7 is a summary of the mysteries that I have come across. All of the data is from my own elicitations and notes unless stated otherwise. For this paper, I have decided to not focus on suprasegmentals, as I do not believe they play any important role in the phonology and it seems that there is no true primary stress unless for emphasizing the importance of a specific word.

2. Consonants

	Bilab.	Labdental	Dental	Alveolar	Post-Alveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b p ^h b ^j		(t) t ^h d d ^j					(k) k ^h g	q		
Nasal	m		n				ɲ	ŋ			
Trill			r								
Tap/Flap			(ɾ)				ɽ				
Fricative		f		s z	ʃ			x	(χ) ʁ		h
Lat. Fric.			ɬ								
Approx.						ɻ	j				
Lat. Approx.			l								

Sounds in parentheses indicate sounds occurring only in allophones of an existing phoneme.

Other phonemes that do not fit in the chart include /w/, /t̂s/, /t̂ʃ/, /t̂ʂʰ/, /t̂ɕ/, /t̂ɕʰ/, and /d̂ʒ/.

2.1 Consonant Environments

Below is a list of the consonant of Balti listed in minimal and near-minimal pairs in the environment before [a], where a word exists (that has been elicited):

1. [p] /pəju/ ‘salt’
2. [p^h] /p^ha/ comparative marker
3. [b] /k^huːrba/ ‘hold’
4. [t^h] /t^hat^ha/ ‘joke’
5. [d] /jidaŋ/ ‘your’
6. [k^h] /k^ha/ ‘upon’
7. [g] /gar/ ‘where’
8. [q] /riχapa/ ‘start’ (Maya Sanchez; 2020/02/14)
9. [m] /ma/ ‘negative marker’
10. [n] /na/ ‘and’
11. [ɲ] /ɲa/ ‘fish’
12. [ŋ] /ŋa/ ‘I’
13. [f] /far/ ‘jump’ (Greta Schatz; 2020/02/18; CM)
14. [s] /sa/ ‘earth’
15. [z] /za/ ‘meal’
16. [ʃ] /ʃa/ ‘meat’
17. [x] /xrta/ ‘horse’
18. [ʁ] /ʁa/ ‘five’
19. [h] /hawu/ ‘wolf’
20. [ʈ] /hʈta/ ‘see’
21. [ɽ] /ɽa/ ‘goat’
22. [j] /jaŋ/ ‘you’
23. [l] /la/ ‘for’
24. [w] /wa/ ‘fox’
25. [t̂s] /t̂sun̂tsɛ/ ‘elder’
26. [t̂ʃ] /t̂ʃulu/ ‘apricot’
27. [t̂ɕ] /t̂ɕa/ ‘tea’
28. [t̂ɕʰ] /t̂ɕʰa/ ‘small grain’
29. [t̂ʂʰ] /t̂ʂʰa/ ‘eagle’
30. [d̂ʒ] /d̂ʒamt̂ʃo/ ‘maternal uncle (respectful)’ (João Costa; 2020/02/04; CM)

2.2 Plosive Symmetry

There appears to exist a symmetry between the plosives, wherein two main types of sounds exist for a place of articulation. voiced unaspirated phoneme, and a voiced palatalized phoneme as well. The first of which is ([p^h], [b]):

31. [p^hiti] ‘bread’
 32. [boŋo] ‘daughter/girl’

The next group comprises ([t^h], [d]):

33. [t^hok^h] ‘roof’
 34. [do] ‘that’

The final group is ([k^h], [g]):

35. [k^ha] ‘upon’
 36. [gans] ‘ice’

There is some debate as to whether the bilabials, alveolars, and velar plosives also have a voiced palatalized version to add to this symmetry; so far I have not seen any minimal pairs containing the palatalized versions of the voiced plosives so I have listed the plosives and [j] as separate phonemes for the sake of continuity (see §7.1):

37. [bjɑ] ‘do’
 38. [gjebrɔl] ‘back-side’

It is important to note that this does not mean that different (unaspirated, not-palatalized) versions of these sounds exist in Balti. The voiceless unaspirated bilabial plosive can often be heard in loanwords, and as listed in Example 1 as well, so it is difficult to tell whether this symmetry actually exists or I am just trying to use for sake of convenience:

39. [pɛn] ‘pen’ (loaned from English)
 40. [pɑkɪstɑn] ‘Pakistan’ (loaned from Urdu)

3. Vowels

	Front Unrounded	Central Unrounded	Back Unrounded	Back Rounded
High	i i: (ɪ)		ʊ	u u:
Mid	e e: (ɛ)	ə		o o:
Low	a a:		(ɑ)	

The true phoneme for /i/ is somewhere in between [i] and [ɪ] but it varies how close it gets to either sound depending on its environment (see Section 5.1), so they can be considered allophones. [u] and [ʊ] are distinguishable, but under certain conditions they are allophones of the same phoneme (see §4).

There are very notable length distinctions, and it is debatable as to whether there are two or three length distinctions. There are at least two for [a] and [o], as will be discussed in the Minimal Pairs section section. [ɑ] is only found in loanwords but it is distinguishable for the speaker, as previously seen in example 39, *pakistan*.

3.1 Vowel Minimal Pairs

There was some debate regarding how many length distinctions there are in Balti, but at least two can be agreed upon between short and long. The following minimal and near-minimal pairs and sentences support the length distinctions and provide evidence for the existence of each vowel:

41. [k^ha] ‘upon’, ‘mouth’
42. [k^ha:] ‘snow’
43. [k^hi] ‘dog’
44. [k^hi:] ‘dog (prepositional object)’
45. [k^he:] ‘snow (prepositional object)’
46. [k^heri] ‘your’
47. [kulɛ] ‘slowly’ (Maya Sanchez; 2020/02/18; GM)
48. [k^hu:rbə] ‘hold’
49. [k^hʊŋ ni] ‘their’
50. [pəju] ‘salt’

The following minimal pairs are questionable, as they are the only ones with three length distinctions (see §6 for more detail):

51. [ts^hõ] ‘lake’
 52. [ʔts^ho] ‘hot’
 53. [ʔts^hoː] ‘grandson’

3.2 Vowel Clusters and Diphthongs

There are rarely vowel clusters and the tendency is for the sounds to diphthongize. However, they do appear very rarely not as a diphthong in a couple words and prepositional phrases which I have clustered as one word:

54. [k^haik^h] ‘some’
 55. [ʔts^hoik^ha] ‘upon the lake’

The examples of diphthongs can be found below:

56. [ft^hʃu^ha] ‘make’
 57. [xr^hts^hoa] ‘grass’

3.3 Vowels and Conjunctions

The only time when a long high front unrounded vowel [iː] occurs is when the preposition marker [i] is combined with the word prior and that word ends in an [i] as well, or when one word ends with [i] and the next starts with [i]. This occurs with the following word combinations:

58. [k^hi i k^ha] /k^hiːk^ha/ ‘upon the dog’
 59. [mɪnmi in] /minmiːn/ ‘will give’
 60. [ʔtʃi in] /ʔtʃiːn/ ‘is’

4. Allophones

There are several examples of allophones in Balti. The alveolar tap, for one, is only seen contrasting in place with the alveolar trill, although difficult to tell which is underlying (see §7.4):

61. [bursɛ] ‘bush’
 62. [rgu] ‘nine’

[e] and [ɛ] also seem to be allophones, as again it is often that the sound seems to be in between the two and varies closer to one or the other in certain situations, both sounding the same to the speaker:

63. [haske] ‘tomorrow’
 64. [sɲɛrma] ‘pepper’

The variation between [ɪ] and [i] has been partially explained in §6.2 but is also recorded here:

65. [ʔɪkʰ] ‘one’
 66. [ɪɲmo] ‘tall’

Although [u] and [ʊ] are separate morphemes, in certain situations they act as allophones, where the speaker did not say there was any distinction in sound:

67. [drun] ‘village’
 68. [dɔxʊn] ‘headscarf’

Free variation occurs with [q] and [χ], as evidenced in the word for ‘hand’, the first elicited in isolation and the second in quicker speech:

69. [laqpo] ‘hand’
 70. [laχpo] ‘hand’

5. Phonotactic Constraints

The allowed syllable types are:

(C)CV
 CCCV
 CVC
 CVCC
 CCVC

Consonant clusters of larger than two consonants (CCCV) only are at the beginning of words, but often the first letter “falls off” in spoken speech:

71. [hrkjɛlba] /rkjɛlba/ ‘to swim’
 72. [bgjɛd] /gjɛd/ ‘eight’

All other syllable patterns can occur anywhere in the word. The most common consonant clusters appear to be a plosive and a glide together, as seen above.

6. Phonological Rules

These are a few of the phonological rules that I have derived from the given data, which are subject to change with more examples. Overall, more vowel rules could be perceived as valid and seem to be more well-represented in the data we have collected thus far, but some plosive rules could be gleaned as well (although difficult to generalize to the natural classes). I have not figured out a proper ordering.

6.1 Vowel Nasalization

Vowels become nasalized when in front of a velar nasal.

$V \rightarrow [+nas] / __ [+cons, +nas, +velar]$

	‘come (cont.)’	‘oily flatbread’	‘look (v.)’	‘yours’
UR	/oŋɛt/	/azok ^h /	/toŋma/	/k ^h jaŋ/
VN	õŋɛt	-	tõŋma	k ^h jãŋ
SR	[õŋɛt]	[azok ^h]	[tõŋma]	[k ^h jãŋ]

6.2 Vowel Tensing

Vowels at the end of a word are tense.

$[+syll, -lax, -lo] \rightarrow [+tense] / __ C_0 \#$

	‘very’	‘two’	‘cat’	‘yesterday’
UR	/ɪʃɪn/	/ɲɪs/	/bɪla/	/gundɛ/
VL	ɪʃɪn	/ɲɪs/	-	gunde
SR	[ɪʃɪn]	[ɲɪs]	[bɪla]	[gunde]

6.3 [d]-Devoicing

[d] is devoiced unless followed by a vowel.

$[d_] \rightarrow [+voice] / __ V$

	‘is’	‘is-QUESTION’
UR	/jəd/	/jəda/
DD	jət	-
SR	[jət]	[jəda]

7. Mysteries

7.1 [j] as its own phoneme

It is still quite unclear as to whether, in general, [j] should be considered its own phoneme all of the time. It certainly is its own phoneme, as seen in *jəd* ('is'), but when succeeding a consonant I am unsure as to whether my marking the [j] as separate is truly the case for each phoneme. I was marking it as part of [d] and [g] because I kept hearing them together in various situations, but I could not find any convincing minimal pairs to support this claim. Perhaps with more data this will become apparent as to whether it is its own sound or two separate sounds, or perhaps for one plosive it is the case and for another it is not that they are different sounds, but for not I still cannot tell for sure.

7.2 Vowel lengths

When eliciting the words 'hot', 'grandson', and 'lake' (as described in §3.6), it turned out that all of them had variations of the length of the [o], which the speaker explained as "very short", "normal", and "long", respectively. However, this length distinction does not exist with any other sounds so I am wondering if perhaps the length distinction that the speaker is perceiving is due to the environment that the word appears in instead.

7.3 Existence of retroflex lateral approximant [ɭ]

I am not entirely sure of whether or not [ɭ] exists. I feel that I may have heard it a couple of times but I do not have enough data to corroborate with to determine whether this was just a mistranscription on about four words or it is a distinctive sound. My current hypothesis is that it is in free variation with [l]. Examples of words I transcribed with [ɭ] are:

- | | |
|---------------------------|-----------|
| 73. [p ^h ɛroɭ] | 'outside' |
| 74. [ɭu] | 'sheep' |

Both of these have been transcribed in different contexts with an [l] instead, but I think I need more examples.

7.4 Rhotics

I cannot figure out which of the rhotics is the "true" phoneme. I believe that [r], [ɾ], and [ɹ] are allophonic but the rule, other than perhaps the [ɹ] sounding most like itself at the beginning of words, is not fully realized. I have been transcribing both sounds using [r] the whole time and only finding a few cases where [ɾ] is used, so it is difficult to have proper evidence to support a sound hypothesis.

8. Conclusion

Although there are still some mysteries with respect to which exact forms alternate with which and how the lengths work themselves out, most of the phonemes and allophones have been successfully determined. More phonological rules will be determined as more sentences are elicited and with more data the rest of the morphemes will most likely be much more well-established as well.