

Historical Phonology and other Observations on Kholosi

Aryaman Arora
Georgetown University

December 6, 2020

Abstract

Kholosi /xolosi/ is an underdocumented Indo-Aryan language spoken only in two villages, Kholus and Gotav, in the Hormozgan province of Iran. It is uniquely situated geographically, entirely surrounded by Iranian-family languages which it is not closely related to. The first English-language scholarly work on the language, Anonby and Bahmani [2016], investigated the Kholosi lexicon. This work presents a grammatical description of Kholosi based on elicitation in 2020 from Ahmed Etebari, a native speaker of Kholosi. Comparisons are made to the other Indo-Aryan languages.

1 Introduction

Kholosi is a thus far little-documented Indo-Aryan language spoken in the villages of Kholus and Gotav in the linguistically diverse province of Hormozgan in Iran, first documented in English-language sources by Anonby and Bahmani [2016]. No comprehensive analysis from the perspective of Indo-Aryan linguistics has been attempted, and further information about the language has remained elusive.

This work delivers a grammatical sketch of Kholosi from a diachronic perspective in the context of the Indo-Aryan language family. For Sanskrit, the IAST transliteration system is used (except with aspiration superscripted for clarity) and this system is extended to other Indo-Aryan languages following Masica [1993]. Standard scholarly Persian transliteration is used.¹

1.1 The setting

Hormozgan province sits on the Persian Gulf, across from the tip of the Arabian peninsula, a strategic chokepoint for trade into the Gulf and thus a linguistic melting pot [Taheri-Ardali, 2017, Anonby and Taheri-Ardali, 2015-2020]. Figure 1 shows the distribution of languages in Hormozgan.

The **Southwestern Iranian** family dominates the Hormozgani linguistic environment. Besides the prestigious variety of Tehrani Persian, this family includes Bandari of Bandar Abbas [Pelevin, 2010], Minabi [Skjærvø, 1975], Keshmi [Anonby, 2016], as well as, to the west, Larestani [Moridi, 2009], Kumzari [Anonby and Yousefian, 2011], and Bashkardi [Skjærvø, 1989]. To the east, the **Northwestern Iranian** languages Balochi [Jahani, 2013] and Koroshi [Nourzaei et al., 2015] are spoken. Finally, on the coast we find the **Semitic** language of Gulf Arabic, as well as Sihhi Arabic on Larak Island [Anonby and Yousefian, 2011].

Kholos itself is a small village of scarcely a thousand residents in western Hormozgan, surrounded by Larestani lects. Nearly all of the inhabitants of the town speak Kholosi and are

¹Thanks to Erik Anonby and his research group for providing some unpublished materials and other invaluable information, and to the instructors of LING-001 at Georgetown, Helen Dominic and Bernie O'Connor, who encouraged me when I wanted to take on this vast project. I also acknowledge several people on Twitter who offered thoughts and theories on some of my work, including Samopriya Basu, Adam Farris, Gopalakrishnan Ramamurthy, Lameen Souag, an anonymous Larestani speaker, among others.

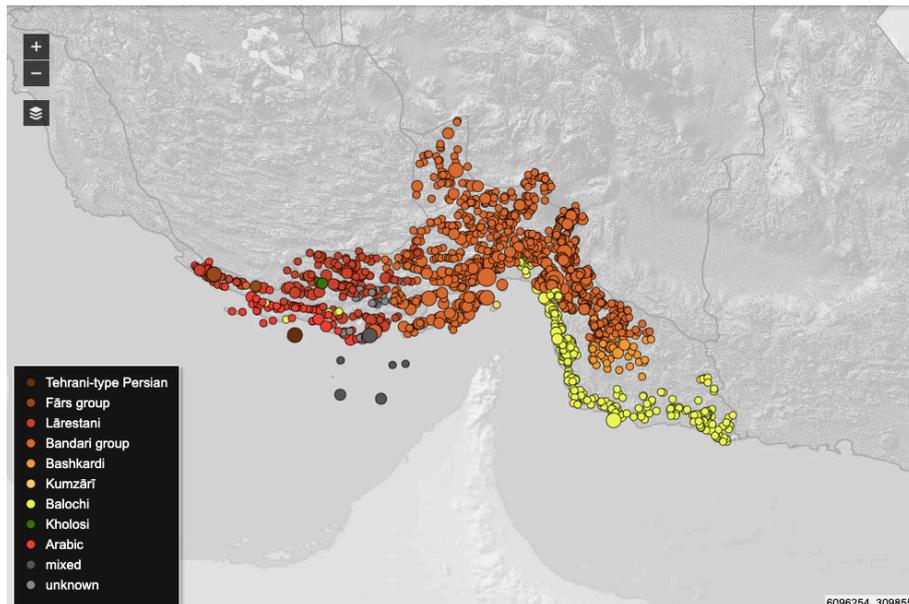


Figure 1: The languages of Hormozgan province, from Anonby and Taheri-Ardali [2015-2020].

Sunni Muslim. My contact, Ahmed Etebari, noted that "Khulus [...] is very deprived in terms of facilities and welfare services, despite being in the best geographical location."

Kholosi is a marginalized lect in Hormozgan. There has been no writing system adopted, no promulgation of a standard, and no official backing by the state. The extremely small number of native speakers, estimated at 1,800 by Anonby and Bahmani (2016), firmly classifies it as an endangered language. However, native speaker attitudes towards the languages do seem positive, and awareness of the language is growing. A field survey would be best suited to assess the sociolinguistic situation in depth, and should be considered in future work.

1.1.1 Indo-Aryan

The geographically closest Indo-Aryan lect to Kholosi is Jadgali in Balochistan, which is as yet undescribed beyond a sociolinguistic survey [Delforooz, 2008]. Hammarström et al. [2020] classifies it as a Sindhic language. Further afield in Iran, we find the nomadic Domari language whose varieties in Jerusalem [Matras, 2012] and Aleppo [Herin, 2012] have been documented, as well as the Zargari lect of the Romani family [Baghbidi, 2003]. Parya, an endangered isolated Central Indo-Aryan language of Tajikistan, has also been documented [Tiwari, 1970].

Anonby and Bahmani [2016] found lexical similarities between Kholosi and the **Sindhic** sub-family of Indo-Aryan. This present work agrees with their findings on Kholosi's Indo-Aryan status, and further explores cross-lingual comparisons with Sindhi [Jetley, 1964, Mewaram, 1910, Trumpp, 1872], the Thari dialect [Bhawnani, 1979], the Jadeja dialect [Mukherjee, 1992], Kutchi (Kacchi) [Rohra, 1966], and Khetrani [Birmani and Ahmed, 2017]. Reference is also made to works on broader Indo-Aryan typology and comparative linguistics [Turner, 1962-1966, Masica, 1993].

1.2 Contact

My language contact is Ahmed Etebari, a 29-year-old native speaker of Kholosi living in Bandar Abbas. His family lives in Kholus. In Bandar Abbas, he owns a business selling masala tea² and promotes the Kholosi language through this venture.

²The local name, probably well-established in the Larestani language, for 'masala tea' is *karak šāy*. This is loaned from Gulf Arabic, and, in turn, *karak* derives from Hindi-Urdu *karak* 'sharp; good, cool'. Otherwise, the standard Persian term for 'masala tea' is *čāy-e māsālā*. Thanks to Alexander Jabbari for this note.

Ahmed is linguistically informed. He is aware of Kholosi’s special status as an effective isolate in Iran, and knows of its lexical similarity to Sindhi. He speaks Persian (both the Larestani dialect around Bastak and standard Iranian Persian) natively, has proficiency in English, and is well-versed in the Classical Arabic *tajwid* reading tradition of the Qur’an.

1.3 Data Collection

All data was collected remotely through WhatsApp audio, in the form of .ogg files.

Phonological analysis was done using the open source software Praat (after conversion to .wav format), as well as a custom JavaScript application using d3.js to visualize and manually classify vowel formants.³ Google Spreadsheets was used to store the lexicon and serve as the backend database for an online Kholosi dictionary page.⁴

I used questionnaires from Abbi [2001] and Anonby and Taheri-Ardali [2020]. The majority of elicitations were produced and organized by myself. My contact further provided natural, unelicited speech and songs in Kholosi.

2 Phonology

2.1 Vowels

	Front	Back	Vowel	F1 (Hz)	F2 (Hz)	Length (ms)
			/i/	330 ± 40	2300 ± 280	170 ± 76
Close	i /i/	u /u/	/ε/	510 ± 70	1990 ± 130	131 ± 74
Mid	e /ε/	o /o/	/a/	710 ± 100	1550 ± 100	140 ± 68
Open	a /a/	ô /ɔ/	/u/ ⁵	370 ± 0	870 ± 0	303 ± 0
			/o/	430 ± 50	910 ± 140	136 ± 60
			/ɔ/	580 ± 50	1020 ± 110	211 ± 69

Table 1: Vowel inventory and key metrics of vowels. Sample standard deviations are provided for numerical measures, with F1 and F2 rounded to the nearest ten.

Kholosi has a six-vowel system, which is minimal for Indo-Aryan; it falls under the *Oriya-type* system of 6 vowels rather than *Marathi–Nepali*, which is remarkable given that Oriya is on the other end of the geographical spread of Indo-Aryan [Masica, 1993, p. 109]. Sindhi and Punjabi have a ten-vowel system that distinguishes vowel length (which Kholosi does not distinguish) and has phonemic /ə/ (which is a possible outcome of an unstressed short vowel in Kholosi) and /e/ (which is allophonic in Kholosi). Typologically, vowel length is quite variable among migratory Indo-Aryan languages with Parya preserving the *Sindhi–Punjabi* 10-vowel system and European Romani varieties having as little as 5 vowels. For all of those languages, Herin [2012]’s description of Domari applies: “more data is needed in order to establish the phonological system”—thus, it is difficult to make any useful comparisons.

The historical sources of the vowels are briefly examined below, but due to the not fully understood system of vowel harmony in Kholosi it is difficult to draw conclusions on historical phonology of the vowels, especially in word-medial positions.

Close front vowel /i/ is well preserved as an outcome of Sanskrit final *-ī*, *-īkā*, *-īya*, *-ī*, and the other sources of the New Indo-Aryan *ī*-stems in Sindhi, e.g. Skt. *pāniya* > *poni* ‘water’, Skt. *mahilā* > *miri* ‘wife’. Some examples of medial development of *i* are Skt. *ślakṣṇa* > *sino* ‘thin’ (cf. Sindhi *sanho*) but Skt. *nikta* > *noko* ‘small’ (cf. Punjabi *nikkā*).

³<https://aryamanarora.github.io/kholosi/vowels>

⁴<https://aryamanarora.github.io/kholosi/dictionary>

⁵Not enough data was sampled for /u/.

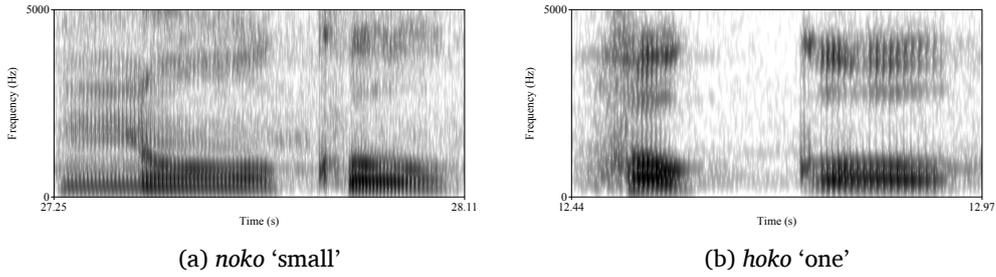


Figure 2: Spectrograms of a near minimal pair, showing vowel length difference for /o/. Also note the difference in voice onset time for the /k/, which is greater in *noko*.

Mid front vowel /ɛ/ is most prominent as part of the infinite suffix *-en* (cf. Sindhi *-aṅ(u)*, Punjabi *-(a)ṅā*, Hindi *-nā*) which derives from the Sanskrit participle in *-anīya*. In syllables with no coda or the coda as a sonorant, /ɛ/ tends to be realised as [e:]. It does not consistently correspond to any single Sanskrit source.

Sometimes it behaves as an epenthetic vowel, e.g. Prs. *abr* > *aber* 'cloud', Skt. *pañca* > *pañjero* 'five'.

Open front vowel /a/ is the outcome of Sanskrit *a*. Like Sindhi, but unlike Hindi, Kholosi does not compensatorily lengthen vowels before geminates, so we find Skt. *sarpa* > Kholosi *sap* 'snake', Sindhi *sapu*, but Hindi *sāp*.

Close back vowel /u/ is the least common vowel. It generally continues Sanskrit *u/ū*, e.g. Skt. *ālu* > *ôlu* 'potato'.

Mid back vowel /o/ continues the masculine ending in Sanskrit *-aka*, which is also the source of *o*-stem nouns in Sindhi and *ā*-stem nouns in Hindi. This includes the masculine endings on declinable adjectives. It also continues Sanskrit *o*, e.g. **koṣma* > *koso* 'warm'.

Open back vowel /ɔ/ continues Sanskrit *ā*. This is actually the reverse development of the Oriya-type six-vowel system, which has Skt. *a* > *ô*, *ā* > *ā̄*.⁶ Some examples are Skt. *māṁsa* > *môz* 'meat', *kṣāra* > *šôr* 'ash'.

2.1.1 Length

No length-based minimal pairs between vowels of the same quality have yet been identified, but some length differences are apparent. E.g. the spectrograms of *noko* 'short' and *hoko* 'one' show variable lengths for /o/ (Figure 2) but this is not a perfect minimal pair. One explanation may be that *hoko* is actually [fiok'.ko], reflecting the gemination in Western Punjabi *hēkk* ~ *hikk* and Punjabi *ikk*. Furthermore, the pitch accent pattern is different between the two words, with *no.ko* being low-falling and *ho.ko* being high-falling.

/ɔ/ tends to be long while the other vowels are more difficult to characterise (Table 1). For example, the minimal pair *maren* 'to die' and *môren* 'to kill' in Figure 3 shows the /ɔ/ in the same position is about twice as long.

2.1.2 Nasalization

A vowel in the vicinity of a nasal consonant is produced with the velum open, a process known as nasalization. This is usual throughout Indo-Aryan. Figure 3 shows nasalization, resulting in the phonetic outcomes [ma.r̥ɛ:n] for *maren* and [mɔ:.r̥ɛ:n] for *môren*.

⁶As Samopriya Basu pointed out, this leads to interesting false friends with Bengali: Kholosi *mar-* 'to die', *môr-* 'to kill' but Bengali *môr-* 'to die', *mār-* 'to kill'.

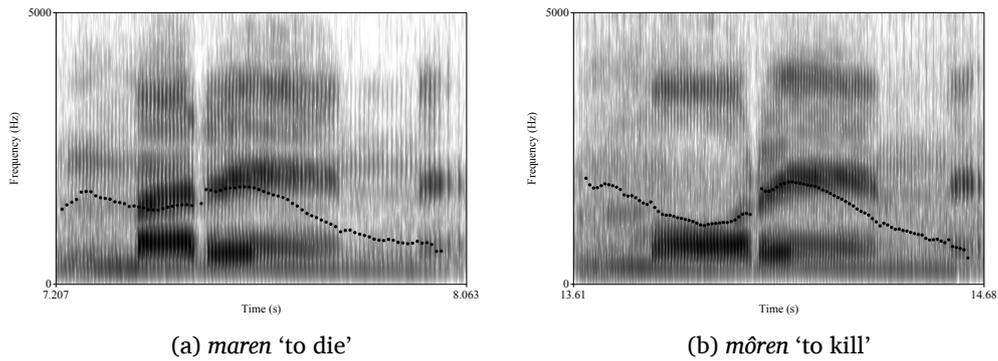


Figure 3: Spectrograms with pitch contour of a minimal pair, showing vowel length difference between /a/ and /ɔ/. Note the nasalization of /ɛ/ partway through production.

2.1.3 Vowel harmony

Vowel harmony is a productive process in Kholosi that is most evident in masculine-feminine alternations caused by the suffixes *-o* (masculine) and *-i* (feminine) in both gendered nouns as well as verbal inflection. Some of the alternations between these two that have been observed are *o~u*, *o/e~i*, and *ô~e*. More data is needed to understand how these processes operate. Typologically, vowel harmony is found in Dardic languages, and, further afield in Indo-Aryan, Bengali and related languages. Persian displays some vowel harmonization but no process as clearly productive as in Kholosi.

/o/ has a lower realisation of [ɔ] when the word is comprised of repeating *Co* units. E.g. /kozoro/ 'man' is realized as [kɔzɔɔɔ].

There were further instances of word-internal vowel changes observed between masculine and feminine forms. It appears that this is result of a broader process of harmonisation of short vowels. This does not appear to be a synchronic process as it was only observed in inherited vocabulary. Some of these forms are listed below.

Sanskrit	Kholosi	Sindhi
<i>ēka</i>	<i>hoko</i> 'one'	<i>hiku</i>
<i>*kuttira</i>	<i>kotoro</i> 'dog'	<i>kuto</i>
<i>markaṭa</i>	<i>moxoro</i> 'ant'	<i>makaṛu</i>
<i>śukrá</i>	<i>sokolo</i> 'white'	–
<i>mūtrá</i>	<i>meter</i> 'urine'	<i>muṭru</i>
<i>putrá</i>	<i>peter</i> 'boy; son'	<i>puṭru</i>

Masculine	Feminine
<i>šorko</i> 'boy'	<i>šurki</i> 'girl'
<i>hoko</i> 'one.M'	<i>hiki</i> 'one.F'
<i>t^hôrgo</i> 'tall.M'	<i>t^hergi</i> 'tall.F'
<i>terero</i> 'three.M'	<i>tiriṛi</i> 'three.F'

2.2 Consonants

	Bilab.	Labiod.	Alveolar	Postalv.	Palatal	Velar	Glottal
Plosive	<i>p</i> /p/		<i>t</i> /t/			<i>k</i> /k/	
	<i>p^h</i> /p ^h /		<i>t^h</i> /t ^h /				
	<i>b</i> /b/		<i>d</i> /d/			<i>g</i> /g/	
Nasal	<i>m</i> /m/		<i>n</i> /n/				
Affricate				<i>č</i> /tʃ/			
				<i>č^h</i> /tʃ ^h /			
				<i>j</i> /dʒ/			
Fricative		<i>f</i> /f/	<i>s</i> /s/	<i>š</i> /ʃ/		<i>x</i> /x/	
			<i>z</i> /z/				<i>h</i> /ɦ/
Tap			<i>r</i> /ɾ/				
Approximant		<i>v</i> /v/	<i>l</i> /l/		<i>y</i> /j/		

Table 2: Consonant inventory.

Voiceless unaspirated plosives /p/, /t/, /k/ constitute this subset of sounds. Synchronically, the phonetic values of /p/ and /k/ are consistent.

Diachronically, /p/ and /k/ continue the same phonemes from Middle Indo-Aryan, while /t/ merges the dental /t̪/ and the retroflex /ʈ/. The assimilation of consonant clusters proceeds as expected in the Sindhic family (and Indo-Aryan in general): stops overpower fricatives which overpower other consonants, and the alveolar stops are the most favoured (e.g. *sapta* > **satta* > *sat-ero* ‘seven’).

One notable feature is the preservation of non-final Indo-Aryan /t̪r/ clusters. Kholosi inserts a vowel in the middle, resulting in /t̪er/, /t̪or/, among others (due to vowel harmony) listed in the table below. Sindhi and Western Punjabi⁷ retain these clusters without epenthesis⁸, but epenthesis occurs in Eastern Punjabi (which has e.g. *putt(ar)* < Skt. *putra*) and the Western Pahari languages [Masica, 1993, p. 200]. Kholosi does not preserve the other *Cr*-clusters (*kr*, *gr*, *ghr*, *pr*, *br*, *bhr*), in line with Sindhi and Punjabi but unlike Dardic.

Phoneme	Sanskrit ⁹	Kholosi
/p/	<i>pānīya</i> <i>sarpa</i>	> <i>poni</i> ‘water’ > <i>sap</i> ‘snake’
/t/	* <i>peṭṭa</i>	> <i>pet</i> ‘belly’
/t/	<i>tārā</i> <i>yukta</i>	> <i>tôro</i> ‘star’ > <i>juti</i> ‘shoes’
/k/	<i>kāla</i> <i>aratnika</i>	> <i>kôro</i> ‘black’ > <i>ôrak</i> ‘elbow’
/t̪er/	<i>putra</i> <i>trayas</i>	> <i>peter</i> ‘son; boy’ > <i>ter-ero</i> ‘three’
/t̪or/	* <i>kutra</i> ¹⁰	> <i>kotoro</i> ‘dog’
/t̪/	<i>rātri</i>	> <i>rôt</i> ‘night’

Table 3: Sources of voiceless unaspirated stops in Kholosi.

Iranian borrowings also have these stops: *parvāz kardan* > *parvôz karen* ‘to fly’, *rāsta* > *rôsta* ‘straight’, *makidan* > *makken* ‘to suck’.

Voiceless aspirated plosives Anonby and Bahmani [2016] were uncertain about the phonemic status of aspiration. In the course of this work, I found a minimal pair distinguished by aspiration: [p^hi.jē:n] ‘to stand’ vs. [pi.jē:n] ‘to drink’. Thus, aspiration is contrastive in Kholosi.

The voiceless aspirates are /p^h/ and /t^h/. The lack of /*k^h/ is noteworthy; it has been lenited to /x/, which is surprising since I do not find any parallel phonemic /f/ or /t/.

These continue the same series from Middle Indo-Aryan (sometimes the result of clusters in Old Indo-Aryan, e.g. *st* > (t)^h) just as Sindhi, Eastern and Western Punjabi, and Gujarati, and indeed the majority of New Indo-Aryan. Like for the unaspirated alveolar stop, the retroflex /ʈ^h/ and dental /t̪^h/ merge into alveolar /t^h/.

The peculiarity in Kholosi is that the Indo-Aryan voiced aspirated (or breathy-voiced) series is devoiced word-initially: /d̪^h/, /d̪^h/ > /t^h/, /b̪^h/ > /p^h/. This has no parallel in Sindhi, but similar processes do occur as we move north through that region, i.e. in Eastern Punjabi, Hindko, several Dardic lects, as well as some Western Pahari languages [Masica, 1993, p. 102]. There is a further division between Punjabi-type languages in which the voiced aspirates develop into tones, and Dardic-type languages which acquired tone through other processes [Baart, 2014].

⁷I refer to what in older scholarship is Lahnda as Western Punjabi, as most modern work does. It should also be noted that Western Punjabi is not a single language but a dialect continuum.

⁸Sindhi actually has /t̪r/ > retroflex /ʈr/ while Western Punjabi keeps the dental, but this difference is irrelevant to Kholosi since it has no retroflex–dental distinction.

⁹All Sanskrit etyma and reconstructed forms are from Turner [1962-1966].

¹⁰I prefer reconstructing this term as **kutra* instead of Turner [1962-1966]’s **kuttira* given the cognates: Gujarati *kuṭrô*, Marathi *kuṭrā*, Bhadravahi *kuṭar*, *kôtar*, *kôtër*, Pangwali *kuttar*, to which I add Chinali *kuṭur* and Gaddi *kuttar*. All reflect a cluster with the usual short-vowel epenthesis in Pahari and none continue *i. Turner’s reconstruction has -i- perhaps to reflect Old Gujarati *kūṭraii* and Western Punjabi *kuṭir* ‘pack of dogs’, but these are hardly enough evidence (especially since the latter appears to have some morphological extension).

Phoneme	Sanskrit	Kholosi
/p ^h /	<i>b^haginī</i>	> <i>p^hēn</i> ‘sister’
	<i>b^hūmi</i>	> <i>p^hi</i> ‘earth’
	<i>*p^hupp^ha</i>	> <i>p^hepi</i> ‘father’s sister’
/t ^h /	<i>dīrg^ha</i> (<i>*d^hīrga</i>)	> <i>t^hōrgo</i> ‘tall, long’
	<i>duhitṛ</i>	> <i>t^hiv</i> ‘girl’
	<i>st^hūla</i>	> <i>t^hōllo</i> ‘fat’

Table 4: Sources of voiceless aspirated stops in Kholosi.

Iranian languages do not provide any loanwords with these phonemes.

Voiced plosives /b/, /d/, and /g/ comprise this class of phonemes. These continue the equivalent Indo-Aryan sources, again with the merger of retroflex /ɖ^h/ and dental /ḍ^h/ into alveolar /d/.

It should be noted that the Indo-Aryan voiced aspirated /g^h/ is not devoiced like the rest of the voiced aspirated series; rather, it is deaspirated or the aspiration is shifted onto an adjacent consonant (e.g. *ghotaka* > **ghoro* > *gohro* ‘horse’). This suggests that the development /k^h/ > /x/ (discussed below) preceded deaspiration, which would render deaspiration an independent development from similar processes in Punjabi and Dardic. Liljegren [2016]’s definitive analysis of Palula, a Dardic language, analyzed the voiced aspirate series as *Ch* clusters; a similar analysis may be fitting for Kholosi to explain the process of aspiration shifting.

Phoneme	Sanskrit	Kholosi
/b/	<i>*bābba</i>	> <i>bōbo</i> ‘father’s brother’
	<i>carvati</i>	> <i>čaben</i> ‘to chew’
/d/	<i>danta</i>	> <i>dand</i> ‘tooth’
	<i>*kaḍḍ^hati</i>	> <i>kadden</i> ‘to pull’
/g/	<i>gāḍ^ha</i>	> <i>gahro</i> ‘red’
	<i>gṛhṇāti</i>	> <i>genen</i> ‘to take, buy’
	<i>maṅgala</i>	> <i>mangal</i> ‘fire’

Table 5: Sources of voiceless aspirated stops in Kholosi.

All of these are well-represented in Iranian loanwords.

Nasals The nasals are /m/ and /n/. Allophonically, /n/ can be realized as [ŋ] and [ɳ] as well. Anonby and Bahmani [2016] posited a retroflex /ɳ/ as a separate nasal phoneme but I was unable to find concrete evidence for this in my data, as two vocabulary terms they found retroflex nasals for had alveolar nasals in my data: *ōno* ‘egg’ and *parono* ‘old’. The retroflex nasal may be subject to speaker variation.

Phoneme	Sanskrit	Kholosi
/m/	<i>medas</i>	> <i>meo</i> ‘fat (noun)’
	<i>āmra</i>	> <i>ambo</i> ‘mango’
/n/	<i>nābhi</i>	> <i>nei</i> ‘navel’
	<i>aṅgula</i>	> <i>ōngul</i> ‘finger’

Table 6: Sources of affricates in Kholosi.

Affricates The affricates are /tʃ/, /tʃ^h/, /dʒ/. Again, these continue the equivalent Indo-Aryan sources.

Some initial /tʃ^h/ have been lenited to /ʃ/, such as *šero* ‘six’ as opposed to Sindhi *c^ha*, *šorko* ‘boy’ for Sindhi *chokro*. It is not clear what conditions this change, but perhaps it is the following vowel, seeing as we only find /tʃ^hi/ in our lexicon.

Phoneme	Sanskrit	Kholosi
/tʃ/	* <i>cavati</i> * <i>gicca</i>	> <i>čayen</i> ‘to tell, say’ > <i>geči</i> ‘neck’
/tʃ ^h /	<i>c^hidati</i> <i>matsya</i>	> <i>č^hiyen</i> ‘to cut’ > <i>mač^hi</i> ‘fish’
/dʒ/	<i>pañca</i>	> <i>panjero</i> ‘five’

Table 7: Sources of affricates in Kholosi.

Fricatives The fricatives are /f/, /ʃ/, /z/, /ʒ/, /x/, and /h/. Of these, all can be derived through native sound changes from Indo-Aryan as well as from Iranian loanwords.

The Indo-Aryan aspirated geminates have collapsed into this fricative series, which is a very unusual change in the context of Indo-Aryan (where aspirates have resisted lenition almost universally). Some Hindko varieties also display frication of this form, adding to evidence that Kholosi has ties to a more northern variety than standard Vicholi Sindhi. Persian contact is also a likely contributor to the loss of these geminates.

The outcome of the Sanskrit thorn cluster *kṣ* is generally *č^h*, which is like that of Marathi, Gujarati, and Sindhi, as opposed to the rest of Indo-Aryan which favours *k^h* except for some specific lexical items that have spread through contact. This is supposedly a distinguishing factor between Outer and Inner Indo-Aryan in some models of Indo-Aryan dialectology; if that grouping is valid, then Kholosi typifies with Outer Indo-Aryan phonologically.

/fi/ is an unstable phoneme, often dropped word-initially (e.g. in the various forms of ‘to be’), following vowels word-internally (e.g. *gahro* ‘red’ realized as /garo/), and intervocalically: *p^hiyen* was recorded for ‘to stand’, but when enunciated clearly again, *p^hihen* was recorded (cf. Sindhi *bihanu*).

- (1) hoko kozoro yu
one.M man be.PST
‘There was a man.’
- (2) môy ahmad i
1SG Ahmad be.PRS.1SG
‘I am Ahmed.’

Tap The only tap is /ɾ/, which is sometimes realised also as a trilled [r]. It also is in free variation with retroflexed /ɽ/, as evidenced by reflexes of Sanskrit *ḍ* which gives a retroflex flap in Sindhi but just /ɾ/ in Kholosi, as well as individual words that were recorded with both phonemes: [dur] ~[dɽ] ‘far’.

Approximants The approximants in Kholosi are /v/, /l/, and /j/. /v/ is variable between [w] (especially between vowels) and [v], while /j/ only ever occurs as an intervocalic glide.

/l/ is an interesting sound, as one of the diagnostic phonemes of Indo-Aryan dialectology. Since the earliest stage of Old Indo-Aryan, there has been difference in how /l/ and /ɾ/ were treated; in some dialects, they merged to one or the other, and in others there was a mix. In Classical Sanskrit, an arbitrary mix of the two was standardized. New Indo-Aryan languages continue to reflect this divide along a very rough Inner-Outer division [Cathcart, 2020]. Kholosi tends towards /l/, certainly more so than Sindhi, as evidenced by examples such as Skt. *maṅgala* > Kholosi *mangal*, Sindhi *maṅgaru* ‘fire’.

Phoneme	Sanskrit	Kholosi
/f/	<i>sarva</i> (* <i>sabbha</i> ?)	> <i>sôf</i> ‘all’
/ʃ/	<i>svapnāyate</i> <i>sidd^ha</i> <i>śyāva</i>	> <i>semen</i> ‘to sleep’ > <i>sida</i> ‘straight’ > <i>sôo</i> ‘green’
/z/	<i>stabd^ha</i> (MIA <i>t^hadd^ha</i>) <i>vaḍra</i> (MIA <i>vaḍḍa</i>) <i>garjara</i> (MIA <i>gājjar</i>) <i>vimśati</i>	> <i>t^hôzo</i> ‘cold’ > <i>vazo</i> ‘big’ > <i>gezar</i> ‘carrot’ > <i>vizero</i> ‘twenty’
/ʒ/	<i>kṣāra</i> (MIA <i>c^hāra</i>) <i>*c^hokara</i>	> <i>šôr</i> ‘ash’ > <i>šorko</i> ‘boy’
/x/	<i>khalla</i> <i>khādati</i>	> <i>xal</i> ‘skin, bark’ > <i>xôyen</i> ‘to eat’
/ɦ/	<i>hasta</i> <i>megha</i>	> <i>hat</i> ‘hand’ > <i>meh</i> ‘rain’

Table 8: Sources of fricatives in Kholosi.

Phoneme	Sanskrit	Kholosi
/ɾ/	<i>rātri</i> <i>guru</i>	> <i>rôt</i> ‘night’ > <i>gaworo</i> ‘heavy’

Table 9: Sources of taps in Kholosi.

Phoneme	Sanskrit	Kholosi
/v/	<i>vyāta</i> , <i>vartman</i> <i>nava</i>	> <i>vôt</i> ‘mouth; path’ > <i>navero</i> ‘nine’
/l/	<i>loman</i> <i>gola</i>	> <i>loy</i> ‘hair’ > <i>golôndo</i> ‘round’
/j/	<i>pībati</i> <i>*d^hauvati</i>	> <i>piyen</i> ‘to drink’ > <i>t^hoyen</i> ‘to wash’

Table 10: Sources of approximants in Kholosi.

2.3 Phonetic processes

2.3.1 Intervocalic lenition

Intervocalic alveolar stops are lenited into fricatives, a phenomenon already observed by Anonby and Bahmani [2016] and tied to similar processes in neighbouring Iranian lects of the Zagros mountain range.

$$\left[\begin{array}{c} + \text{stop} \\ + \text{alveolar} \end{array} \right] \rightarrow \left[+ \text{fricative} \right] / \text{V_V}$$

This may not be an obligatory process however. At least in the elicitation of individual words, there was variation in the occurrence of this phenomenon, as seen in Figure 4.

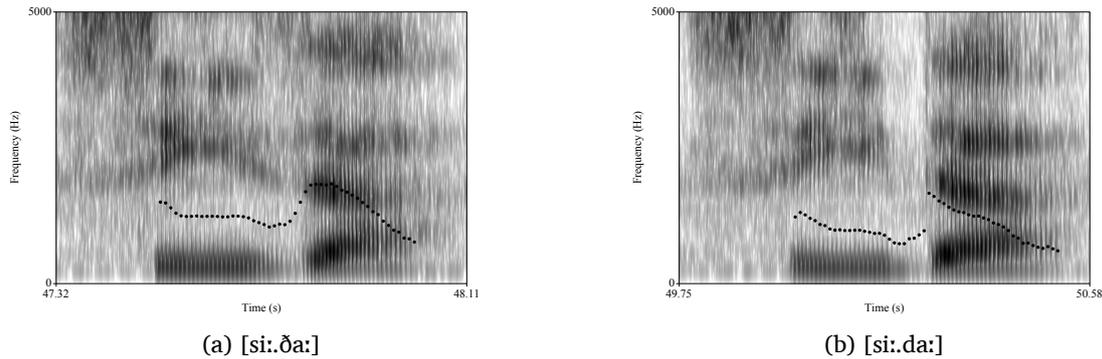


Figure 4: Two realizations of *sida* ‘straight’, the first showing intervocalic lenition while the second shows the plain stop.

2.3.2 Pitch accent

Pitch accent is undoubtedly one of the most interesting phonological characteristics of Kholosi. As in Anonby and Bahmani [2016], I found strong evidence that pitch is important for word and syllable demarcation, if not lexically contrastive. Vowel length also is tied to pitch; non-final syllables with long vowel nuclei tend to take a low pitch, as in *šorko*. Figure 5 shows some of the different pitch contours recorded.

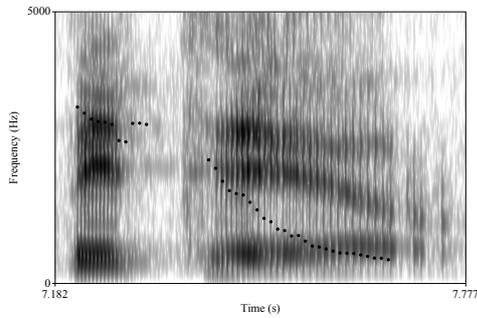
Anonby mentioned a potential relation between morphological class and pitch accent, and I found this exact connection in numerals, which all bear high pitch on the first syllable which progressively falls over the word.

3 Morphology

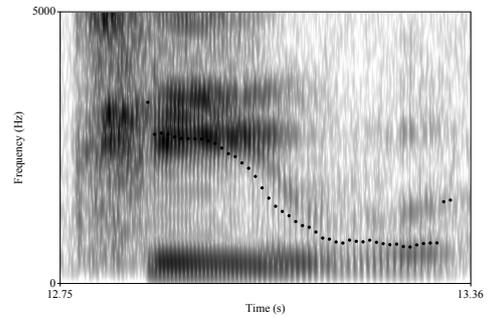
Work on Kholosi morphology is underway. Some of the interesting features are discussed below.

Gender Kholosi has two grammatical genders: masculine and feminine. This is in line with Sindhi, Punjabi, and Hindi, and unlike the three-gender system of Gujarati and Marathi. Not all adjectives are declinable, but those that are alternate between the gender morphemes *-o* (M) and *-i* (F).

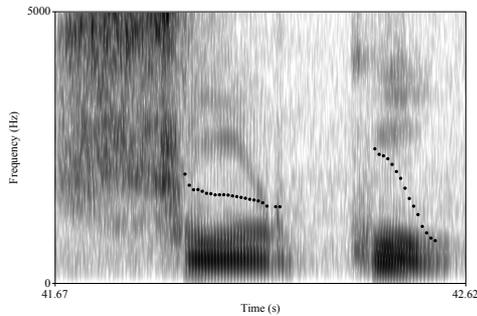
Some nouns show clear gender alternations (often with vowel harmony), e.g. *nôno* ‘maternal grandfather’ vs. *noni* ‘maternal grandfather’.



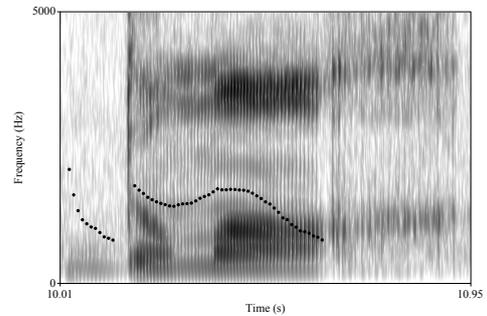
(a) *pe.ter* [pe.tɛ:r] H.HL 'son'



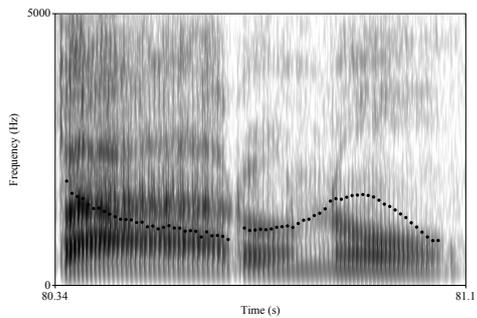
(b) *t'iv* [tʰiv] HL 'daughter'



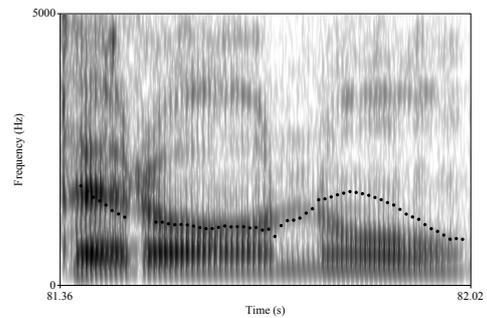
(c) *šor.ko* [ʃo:r.ko] L.HL 'boy'



(d) *do.môx* [do.mɔ:x] M.HL? 'nose'



(e) *pa.ro.no* [pa:ro.no] HL.L.HL 'old'



(f) *pe.rô.no* [pe.rɔ: no] HL.L.HL 'old'

Figure 5: Spectrograms with pitch contours for several words. Note the visible aspiration at the start of *t'iv* and the variable vowel lengths in each word. The identical pitch accent of *parono* and *perôno* is remarkable; they appear to be vowel-length differentiated variants of the same word, a mechanism that is as yet not understood.

1	<i>hoko ~ hiki</i>	11	<i>dazohoko</i>
2	<i>baro</i>	12	<i>dazobaro</i>
3	<i>terero ~ tiriṛi</i>		...
4	<i>čôllo</i>		
5	<i>panjero ~ panjiri</i>	20	<i>vizero</i>
6	<i>šero</i>	21	<i>vizohoko</i>
7	<i>satero</i>		...
8	<i>atero</i>		
9	<i>navero</i>		
10	<i>dazero</i>	100	<i>sau</i>

Table 11: Numerals

Numerals Only numerals 1–29 and 100 were known to my contact, suggesting large-scale adoption of Persian. My contact, for example, used Persian numerals in the phrases ‘two weeks’ (*do hafta*) and ‘one week’ (*yak hafta*).

Every numeral term has the declinable suffix *-ro ~ -ri* attached, which is dropped when it is in the tens place. This may be related to the higher numeral suffix *-ro* in Vicholi Sindhi and *-ḍo* of *ekḍo* ‘one’ in the Jadeja dialect [Trumpp, 1872, Mukherjee, 1992].

The structure of *hoko* ‘one’ aligns with Western Punjabi *hekk/hikk* and Sindhi *heko*, which are all rooted in Sanskrit *eka*. Berger [1992] suggests that the gemination (also encountered in Middle Indo-Aryan *ekka*) and the initial *h-* is of emphatic origin.

Pronouns The pronouns are:

	1	2	3.PROX	3.DIST
SG	<i>môy</i>	<i>attu</i>	<i>he</i>	<i>ho</i>
PL	<i>ôse</i>	<i>ôve</i>	<i>hoven</i>	<i>hozen</i>

The proximal vs. distal distinction for third person, lack of gender distinction, and the forms of the pronouns themselves are not unusual for Indo-Aryan or the Sindhic subfamily.

Vicholi Sindhi actually has *āūm* for 1SG, while northern dialects have *mān* like Kholosi.

The 2PL of Kholosi is unusual, with possible cognates only found in Khetrani *avhe*, Lasi Sindhi *avī*, and Sinhala *u^mbə* (which is 2SG), ultimately from Sanskrit (*y*)*uṣmad*.

Pronominal suffixes Like Sindhi and Persian, Kholosi has genitive pronominal suffixes that attach to the possessum, as well as the usual Indo-Aryan genitive suffix (*jo* in Kholosi, just as in Sindhi) if the possessor is not a pronoun.

- (3) *môy gohr = oy* *kôr-o* *he*
 1SG horse.M = 1SG black-M COP.PRS.3SG
 ‘My horse is black.’

Type	Kholosi	Sindhi	Persian
1SG	<i>-oy</i>	<i>-am</i>	<i>-am</i>
2SG	<i>-o</i>	<i>-aṇ</i>	<i>-at</i>
3SG	<i>-os</i>	<i>-as</i>	<i>-aš</i>
1PL	<i>-om</i>	—	<i>-emān</i>
1PL	<i>-om</i>	—	<i>-etān</i>
1PL	<i>-ôn</i>	—	<i>-ešān</i>

Table 12: Pronominal Suffixes

Complementizer The complementizer *jo* (not to be confused with the identical genitive suffix) is very prominent in spoken Kholosi. It likely derives from Sanskrit *yah*, a reflexive pronoun, which it also sometimes behaves as.

- (4) hok-o kozoro hu **jo** nôy = os ali hu
 one-M man.M COP.PST.3SG REL name = 3SG Ali COP.PST.3SG
 ‘There was a man **whose** name was Ali.’
- (5) deh **jo** hamo zendegi kere-may
 village.M REL 1PL.ERG? life do-PRS
 ‘the village **that** we live in’

Case markers Kholosi has a full case-marking system like most Indo-Aryan languages of the region. The accusative–dative merger is widespread in Indo-Aryan and is reflected here as well. Notably, Kholosi allows marking both the accusative and dative at the same time (even though both cases have the same marker, *ke*), unlike Hindi which only marks dative if both semantic roles are present in the frame. Kholosi also merges allative and instrumental.

Function	Kholosi	Etymology
ACC/DAT	<i>ke</i>	Skt. <i>kr̥te</i>
LOC	<i>mô</i>	Skt. <i>madhya</i>
GEN	<i>jo</i>	cf. Sindhi <i>jo</i>
COM	<i>sônda</i>	Skt. <i>santaka</i>
ALL/INS	<i>te</i>	cf. Punjabi <i>utte</i> , <i>te</i> ‘on’
ABL	<i>tôw</i>	cf. Punjabi <i>tom</i>

Table 13: Case markers

Compound verbs An areal feature of both Iranian and Indo-Aryan, as well as Turkic and Uralic languages, is the use of compound verbs made up of noun–verb or verb–verb concatenations. Kholosi is no exception, with much of the basic vocabulary for verbs being compounds. Some examples were *čarx diyen* ‘to turn’ (lit. ‘to give wheel’), *čuri karen* ‘to steal’ (lit. ‘to do theft’), *feker karen* ‘to think’ (lit. ‘to do thinking’, calqued from Prs. *fekr kardan*), *gap č^hiyen* ‘to talk, converse’ (lit. ‘to hit talk’, calqued from Prs. *gap zadan*).

4 Lexicon

I collected 332 Kholosi words, of which 46% were identified to be of Sanskrit origin while 29% were Persian (or Larestani) borrowings. There is some contact with Arabic (1%), English (1%), and Balochi (possibly one word, *kend* ‘knee’ < Balochi *kóndh*). 18% were of unknown origin, and are likely mostly Larestani dialectal borrowings that were difficult to identify without access to a speaker or linguistic materials. The rest were synchronic derivations.

5 Future work

There is an ongoing effort on Kholosi grammatical description and documentation underway by Maryam Nourzaei at Uppsala University, which I eagerly await in order to have better synchronic description of the language. I hope future descriptive work can shed light on vowel harmony, the Kholosi verbal system, and codeswitching with Persian.

Much investigation is still necessary on the diachronic end. This work has shown that Kholosi has some tantalizing similarities to lects north of Sindhi, but ultimately confirms its status as a Sindhic subfamily lect. There is also work to be done from the sociolinguistic perspective given the unique contact between Indo-Aryan and Iranian taking place in Kholosi, and the resilience of the language in spite of its miniscule speaker population.

Kholosi is small, but in no danger of extinction given what my contact has told me about the continuing use of the language. It will be exciting to see Kholosi have an Internet presence, which is why I have released my fieldwork data online at <https://aryamanarora.github.io/kholosi/>.

*mohemtarin kôm môrez = jo mehrabôni he
egah koi môn = tôw komak konjai = yā
be he mani he jo
môy hanuzô p^{ti} mate arzeš hati*

Kindness is the most important principle of humanity.
If anyone asks me for help
This means that,
I'm still valuable on earth.
—Ahmed Etebari

References

- Anvita Abbi. *Basic sentences for collecting field data in Indian languages*. LINCOM, 2001.
- Erik Anonby. The Keshmi (Qeshmi) dialect of Hormozgan province, Iran: A first account. *Studia Iranica*, 44(2):165–206, 2016.
- Erik Anonby and Hassan Mohebbi Bahmani. Shipwrecked and landlocked: Kholosi, an Indo-Aryan language in south-west Iran. *Studia Iranica*, 43:1–23, 2016.
- Erik Anonby and Mortaza Taheri-Ardali, editors. *Atlas of the Languages of Iran*. Geomatics and Cartographic Research Centre, 2015-2020.
- Erik Anonby and Mortaza Taheri-Ardali, editors. *Language data questionnaire*. In Anonby and Taheri-Ardali [2015-2020], April 2020. URL <https://carleton.ca/iran/wp-content/uploads/ALI-Qaire-EN-20200402.pdf>.
- Erik Anonby and Pakzad Yousefian. *Adaptive multilinguals: A survey of language on Larak Island*. Uppsala University, 2011.
- Joan Baart. Tone and stress in North-West Indo-Aryan: A survey. In *Above and Beyond the Segments*, pages 1–13. John Benjamins, 2014. URL <https://www.jbe-platform.com/content/books/9789027269065-z.189.01baa>.
- Hassan Rezai Baghbidi. The Zargari language: An endangered European Romani in Iran. *Romani Studies*, 13(2):123, 2003.
- Hermann Berger. *Modern Indo-Aryan*, pages 243–288. Mouton de Gruyter, 1992.
- Murli D. Bhawnani. *Descriptive analysis of Thari: A dialect of Sindhi language*. PhD thesis, 1979.
- Ali H. Birmani and Fasih Ahmed. Language of the Khetrans of Barkhan of Pakistani Balochistan: A preliminary description. *Lingua*, 191:3–21, 2017.
- Chundra A. Cathcart. A probabilistic assessment of the indo-aryan inner–outer hypothesis. *Journal of Historical Linguistics*, 10(1):42–86, 2020. ISSN 2210-2116. doi: <https://doi.org/10.1075/jhl.18038.cat>. URL <https://www.jbe-platform.com/content/journals/10.1075/jhl.18038.cat>.
- Behrooz Barjasteh Delforoos. A sociolinguistic survey among the Jadgal in Iranian Balochistan. In Carina Jahani, Agnes Korn, and Paul Brian Titus, editors, *The Baloch and Others: Linguistic, Historical and Socio-political Perspectives on Pluralism in Balochistan*, pages 23–43. Reichert, Wiesbaden, 2008.
- Harald Hammarström, Robert Forkel, Martin Haspelmath, and Sebastian Bank. Glottolog 4.3, 2020. URL <https://glottolog.org/>.
- Bruno Herin. The Domari language of Aleppo (Syria). *Linguistic Discovery*, 10(2), 2012.
- Carina Jahani. The Balochi language and languages in Iranian Balochistan. *The Journal of the Middle East and Africa*, 4(2):153–167, 2013.
- Murlidhar Kishinchand Jetley. *Morphology of Sindhi: A Descriptive Analysis of Vicoli, the Standard Sindhi Dialect*. PhD thesis, 1964.
- Henrik Liljegren. *A grammar of Palula*. Language Science Press, 2016. URL <https://library.oapen.org/bitstream/handle/20.500.12657/32374/611690.pdf?sequence=1>.
- Colin P. Masica. *The Indo-Aryan languages*. Cambridge Language Surveys. Cambridge University Press, Cambridge, 1993. ISBN 9780521299442.
- Yaron Matras. *A Grammar of Domari*. De Gruyter, 2012. ISBN 9783110291421.

- Parmanand Mewaram. *A Sindhi-English dictionary*. The Sind Juvenile Co-operative Society, 1910. URL <https://dsal.uchicago.edu/dictionaries/mewaram/>.
- Behzad Moridi. The dialects of Lar (the state of research). *Iran the Caucasus*, 13(2):335–340, 2009. ISSN 16098498. URL <http://www.jstor.org/stable/25703812>.
- Sibasis Mukherjee. *Phonology and morphology of Jadeja dialect of Sindhi*. PhD thesis, 1992.
- Maryam Nourzaei, Carina Jahani, Erik Anonby, and Abbas Ali Ahangar. *Koroshi: A corpus-based grammatical description*. Acta Universitatis Upsaliensis, 2015.
- Mikhail Pelevin. Materials on the Bandari dialect. *Iran the Caucasus*, 14(1):57–78, 2010. ISSN 16098498. URL <http://www.jstor.org/stable/25703833>.
- Satish Kumar Rohra. *Descriptive analysis of Kachi (Kacchi)*. PhD thesis, Poona: Deccan College, 1966.
- Prods O. Skjærvø. Notes on the dialects of Minab and Hormoz. *Norwegian Journal of Linguistics*, 29:113–128, 1975.
- Prods O. Skjærvø. Languages of Southeast Iran: Lārestānī, Kumzāri, Baškardī. In Rüdiger Schmitt, editor, *Compendium Linguarum Iranicarum*, pages 363–369. Dr. Ludwig Reichert, Wiesbaden, 1989.
- Mortaza Taheri-Ardali. A first account of the languages in western Hormozgan. August 2017.
- Bholanath Tiwari. *Tajuzbeki*. National Publishing House, 1970. URL <https://archive.org/details/Tajuzbeki-Dr.BholanathTiwari>.
- Ernest Trumpp. *Grammar of the Sindhi Language*. Trübner, London, 1872.
- Ralph Lilley Turner. *A comparative dictionary of Indo-Aryan languages*. Oxford University Press, 1962-1966. URL <https://dsal.srv04.uchicago.edu/dictionaries/soas/>.