



The
Word Order in
Western Asia
Corpus

metadata for the data set

Iranian — Kumzari (Musandam)

24 August 2021
(doc.v1.0)

Overview of the data set

data set	iran_kumzari_musandam
contributors	Geoffrey Haig
published	24 August 2021
texts	2
tokens	592 analyzed (1002 total)
notes	XLS/TSV contain entire source text

Metadata for individual texts

Text A

source	Anonby (2015: 262–273) (WAV ✗ PDF ✓)
location	Kumzar, Musandam, Oman (26.3375° N, 56.4099° E)
recorded	c. 2010
text type	traditional narrative
speaker	01 — gender unknown, age unknown
tokens	416 analyzed (560 total)

Text B

source	Anonby (2015: 278–286) (WAV ✗ PDF ✓)
location	Kumzar, Musandam, Oman (26.3375° N, 56.4099° E)
recorded	c. 2010
text type	traditional narrative
speaker	02 — gender unknown, age unknown
tokens	176 analyzed (442 total)

Background to the data set

The texts in this data set, and the information necessary for analysis, has been taken from the rich documentation of Anonby (2015), which is fully available online.¹

The Kumzari speech community live in isolated locations “on the Musandam peninsula of northern Oman. They number approximately 5 000, including 500 living across the Strait of Hormuz on Larak Island, Iran and several families across the border in the United Arab Emirates. Semi-nomadic, they migrate between the winter coastal villages and oasis towns such as Khasab and Dibba in summer. Their geographic focal point is the village of Kumzar, situated at 26.3375° N, 56.4099° E, and is the northernmost settlement in Oman” (Anonby 2015: 1).

According to Anonby (2015: 3), Kumzari is a language of mixed Iranian and Semitic ancestry. Major aspects of morphosyntax can be assigned to either language family, while the basic lexicon, as far as is known, appears to be about 50% of Iranian origin. While the sheer bulk of Semitic influence in phonology, lexicon, and aspects of morphosyntax is undeniable, there are nonetheless traces of basic Iranian morphology which are less readily reconciled with a mixed scenario, including pronominal paradigms, person and number indexing morphology, suffixal expressions of indefiniteness and definiteness, prefixal aspectual morphology on verbs, and a number of basic verb roots, some of which include irregular secondary stems with clear Iranian cognates (e.g. *kaft*, *kō* ‘fall’; *gaft*, *gō* ‘say’; *gir* ‘take’; *xōr* ‘eat’; *bzēn* ‘hit’; *škašt*, *škēn* ‘break’; *raft*, *čō*, *rō* ‘go’; Anonby 2015: 92). Furthermore, the consistent OV word order with nominal (but not pronominal) direct objects, and the typically Iranian placement of pronominal objects immediately following (in Iranian in fact cliticized to) the non-verbal complement of a light verb likewise bespeak of Iranian roots.

The issues here go far beyond what can be accommodated in this brief metadata sketch; we will simply note that for the practical purposes of the present project, we have opted to include Kumzari within the Iranian languages. Note that this is not intended as a classification of the language in terms of its current structure, but as an approximation of its historical origins. It can be compared to, for example, Matras’ (2020: 21) characterization of Jerusalem Domari as a language of ‘Indo-Aryan origin’, despite the immense overlay of Arabic influence on this language. Although our decision is clearly problematic, it appears somewhat less so than classifying Kumzari as a Semitic language, currently the only other option available to us. It is also the decision which has been most widely favoured by other scholars consulted on this. Ultimately, we hope that the research agenda of this project may be able to contribute to a better understanding of the history of Kumzari.

The texts from Anonby (2015) are ‘from recorded and transcribed folktales and from field notes’, and illustrate the Kumzari tradition of oral story-telling. Information on the speakers and their background is unfortunately not available. Text A, *Pačaxčēō*, contains a large number of utterances that had to be left unanalysed due to uncertainties of interpretation, and was therefore not completely analysed. The second text, *Ĝrābō*, yields sufficient analyzable tokens for the data set.

References

- Anonby, Christina. 2015. *A grammar of Kumzari: A mixed Perso-Arabian language of Oman*. Ph.D. dissertation, Leiden University. (<http://hdl.handle.net/1887/32793>).
- Matras, Yaron. 2020. Jerusalem Domari. In Lucas, Christopher & Manfredi, Stefano (eds.), *Arabic and contact-induced change*, 511–531. Berlin: Language Science Press.

multicast.aspra.uni-bamberg.de/resources/wowa/

¹ <http://hdl.handle.net/1887/32793>