

Corpus description for:

Belelli, Sara. 2021. *The Laki variety of Harsin: Corpus and sound files*.
(multicast.aspra.uni-bamberg.de/resources/kurdish/#laki)

Overview of the Harsini corpus

This corpus contains a selection of seven texts recorded between the 10th of January 2014 (20th of Dey 1392) and the 27th of February 2014 (8th of Esfand 1392) in the city of Harsin. Each text has been given a title, which was either provided by the narrator at the moment of recording, or chosen arbitrarily on the basis of the narrative's content. A synopsis of the seven texts with basic metadata on the speakers is given in the Table 2 at the end of this document; page numbers refer to Belelli (in print), which should be consulted for details of the analysis.

The arrangement of the seven texts follows a standardised format. Each of them is preceded by a few introductory paragraphs containing relevant information on the narrator and the recording situation. The introduction is followed by a summary of the narrative's content, containing also basic information on the folktale's typology.

The texts in transcription and translation have been divided into numbered paragraphs or annotation units, structured as follows:

[1:12] S.S. *kwirdî harsîni!* Harsini Kurdish! (00:20)

The segmentation was mainly based on semantic cues and speech pauses. Each unit is identified by two numbers in square brackets on the left side, respectively indicating the text number and the paragraph number: e.g. [1:12]. Only in Text 1, where more than one consultant intervenes in the narration, the speaker's initials have been indicated for each utterance: e.g. S.S. Every annotation has a univocal correspondence with a section of the audio file, indicated by a reference to the beginning time expressed in minutes and seconds: e.g. (00:20).

Transcription conventions are summarized in Table 1 at the beginning of Belelli (in print, p. ix). The adopted representation is broadly phonemic, with some instances of phonetic rendering, which have been signalled and motivated in the sections on Harsini phonetics and phonology (Belelli, in print, Chapter 2, pp. 39-70). Unresolved problems concerning certain sounds might have caused the persistence of incongruities, which will hopefully be clarified and amended by future research.

The transcription of Harsini texts is paralleled by a free English translation. In the English version, the primary intent was to preserve a clear correspondence between word and meaning. Punctuation largely follows the standards of written English. In order to help the reader identify correspondences between the Harsini transcription and the English translation, punctuation has also been added to the transcribed text. The punctuation marks in transcription and translation correspond for the most part, but occasional differences might be observed. Further remarks on form or content are contained in footnotes.

I have tried to represent speech pauses, errors, fragments and hesitations as accurately as possible, in order to give an idea of the complex style of the elocution. For this purpose, the conventional symbols in Table 1 have been used:

Table 1: Symbols used in transcriptions and translations

{ }	Curly brackets include not-analyzed language material, usually representing instances of code-switching to Persian or utterances pronounced by non-native speakers, both in transcription and in translation. These do not include integrated loans or occasional lexical borrowings.
{xxx}	The symbol ‘xxx’ included in curly brackets substitutes linguistic strings that are not clear in sense or form, both in transcription and in translation.
{? }	Uncertain passages, whose transcription or interpretation is provisional, are included in curly brackets and preceded by a question mark.
{# }	Possible speech errors, hesitations and speech fragments are preceded by an asterisk and included in curly brackets, both in transcription and in translation.
[...]	Three dots in square brackets might indicate a silence, an intentional omission or an interruption in discourse. Whenever present, these phenomena are signalled by the same symbol in translation.
[]	Square brackets including language material are used for several reasons. In transcription, they occur whenever the speaker addresses the listener directly, uttering sentences that have little or nothing to do with the narrative plot. In translation, they contain clarifications, preceded by the abbreviation ‘i.e.’ in the case of unspecified referents, or by ‘lit.’ in the case of necessary literal rephrasing, explaining idioms or other unusual forms in English.
()	Parentheses are used to enclose missing language material or unpronounced parts of words in transcription. In translation, they identify the supplemental linguistic material added to the English rendering in order to make the translation comprehensible.

Table 2: Synopsis of the Harsini texts

Text	Speaker	Sex	Age	Genre	min:sec	Book pages
1. <i>Nameyna</i> (The felt-made girl)	M.G./ S.S.	F/ F	50/ 62	Tale of magic (version 1)	10:00	pp. 231- 252
2. <i>Čuwīna</i> (The wooden girl)	F.P.	F	46	Tale of magic (version 2)	10:11	pp. 253- 271
3. <i>Šū kačať</i> (The bald groom)	S.S.	F	62	Tale of magic	04:45	pp. 273- 282
4. <i>Mā(h) pīšānī</i> (Moon-forehead)	S.S.	F	62	Tale of magic	09:30	pp. 283- 300
5. <i>Mardi</i> <i>xīyāťpardāz</i> (The dreaming man)	M.J.	M	50	Anecdote/joke	04:26	pp. 301- 307
6. <i>Žin līva</i> (The crazy wife)	M.J.	M	50	Anecdote/joke	14:17	pp. 309- 328
7. <i>Mīmī nařadō</i> (The demonic aunt)	M.G.	F	50	Tale of magic	05:44	pp. 329- 341
Total					58:53	

References: Belelli, Sara. in print. *The Laki variety of Harsin: Grammar, texts, lexicon*. Bamberg: University of Bamberg Press.