

Prediction Experiment on Western Kho-Bwa languages

Timotheus A. Bodt (SOAS, London) and Johann-Mattis List (Max Planck Institute, Jena)

This database includes all the sound files and the transcriptions of the prediction experiment for Western Kho-Bwa. This experiment was registered as:

Bodt, Timotheus A., Nathan W. Hill and Johann-Mattis List. 2018. *Prediction experiment for missing words in Kho-Bwa language data*. Open Science Framework Preregistrations October 5. <https://osf.io/evcbp/>

The data and code can be found on:

Timotheus A. Bodt, Nathan W. Hill, & Johann-Mattis List. (2018, October 8). Prediction experiment for missing words in Kho-Bwa language data (Version v1.0.1). Zenodo. <http://doi.org/10.5281/zenodo.1451176>

A paper explaining the experiment is under review:

Bodt, Timotheus A. and Johann-Mattis List. 2019 (under review). Testing the predictive force of the comparative method: An ongoing experiment on unattested words in Western Kho-Bwa languages. *Papers in Historical Phonology* Volume 1: 1–21.

The results of the experiment will be presented at the International Conference on Historical Linguistics 24: 01-Jul-2019 - 05-Jul-2019, Canberra, Australia.

The uncut sound files, cut sound files, original field notes and preliminary transcriptions have been saved as:

Bodt, Timotheus Adrianus. (2019). *'Retrodiction' experiment Western Kho-Bwa languages: data [Data set]*. Zenodo. <http://doi.org/10.5281/zenodo.2529727>

How to use these files?

- Download the zip folder soundfiles_prediction_experiment.zip
- Extract the files in a separate folder
- Search for the required sound file(s)

Searching sound files can best be done using the English CONCEPTS from the predictions_results.csv file. For example, searching for BACK will give all the sound files that contain the English gloss 'back' (including 'backwards', 'back' as body part, turn 'back' etc.).

Another option is the select all the sound files of a given linguistic variety / doculect by searching for the original sound file number (i.e. searching for SART111118 will give all the sound files for the Khoitam variety of Sartang):

File name	Language	Variety
SART261018A	Sartang	Khoina
SART261018B	Sartang	Khoina

SART261018C	Sartang	Jerigaon
SART111118	Sartang	Khoitam
SART121118	Sartang	Rahung
SHER251018A	Sherdukpen	Shergaon
SHER251018B	Sherdukpen	Shergaon
SHER101118A	Sherdukpen	Rupa
SHER101118B	Sherdukpen	Rupa
LISH041118	Khispi	Lis
CHUK130115D	Duhumbi	Tsangpa
CHUK090314A	Duhumbi	Duhum
CHUK090314A	Duhumbi	Duhum

I would advise against using a certain attested form in the predictions_results.csv file and search for that (e.g. p a η + b u ‘chest’), because the cut sound files have been saved without spaces and morpheme breaks and because the actual transcriptions of the sound files may have changed after analysis, but were not updated in the name of the cut sound files.

If you cannot find a certain sound file, then it may simply not have been recorded or not cut from the main sound file. If you are really interested, please mail me at timintibet@hotmail.com and I will attempt to find it or record it.

The metadata of the speakers are provided in the following table (age of respondent at the time of the interview):

File name	Date	Name	Language	Variety	Sex	Age
SART261018A	26102018	Phinje Nasidu	Sartang	Khoina	F	34
SART261018B	26102018	Phinje Nasidu	Sartang	Khoina	F	34
SART261018C	26102018	Pema Choijom Yamnojee	Sartang	Jerigaon	F	27
SART111118	11112018	Kezang Rokpu	Sartang	Khoitam	M	57
SART121118	12112018	Chomu Sarmu	Sartang	Rahung	F	27
SHER251018A	25102018	Dombu Tsering Thongon Lama	Sherdukpen	Shergaon	M	71
SHER251018B	25102018	Dombu Tsering Thongon Lama	Sherdukpen	Shergaon	M	71
SHER101118A	10112018	Pema Sinchaji	Sherdukpen	Rupa	M	76
SHER101118B	10112018	Pema Sinchaji	Sherdukpen	Rupa	M	76
CHUK090314A	09032014	Rinchen Butri	Duhumbi	Duhum	F	50
CHUK090314B	09032014	Rinchen Butri	Duhumbi	Duhum	F	50
CHUK130115D	13012015	Dorji Choijom	Duhumbi	Tsangpa	F	43
LISH041118	04112018	Norbu Dema	Khispi	Lish	F	39