# Corpus-based study of variation and language change in Kullui

Anastasia Krylova[2], Julia Mazurova[1], Evgeniya Renkovskaya[1,2]

*1. Institute of Linguistics, Russian Academy of Sciences*

*2. Institute of Oriental Studies, Russian Academy of Sciences*

SOUTH ASIAN LANGUAGES ANALYSIS ROUNDTABLE (SALA-35)
INALCO-PARIS, OCTOBER 29-31, 2019

# Kullu District (Census 2011)

| | |
|---|---|
| TOTAL AREA . . . . . | 5503 Sq.Kms |
| NUMBER OF TAHSIL/ SUB-TAHSILS' . . . . . | 6 |
| NUMBER OF TOWNS . . . | 5 |
| NUMBER OF VILLAGES. . | 326 |

- Population 437 474

# Project on documentation of Kullui

- Launched in 2014 in the Institute of Linguistics, Russian Academy of Sciences
- Funding:
  - ***2014-2018 Russian Foundation for Basic Research (RFBR) № 16-34-01040 «Grammar description and lexicon of Kullui»***
  - ***2019-2021 Russian Foundation for Basic Research (RFBR) № 19-012-00355 «The investigation of the Kullui language based on the oral text corpus»***
- Methods - linguistic fieldwork
- The main objective of the project is documentation of the Kullui language
- Four linguistic expeditions to Kullu district – in 2014, 2016, 2017 and 2018.

# Why Kullui?

- Kullui as a starting point for the description of other Himachali Pahari languages
- Definitely endangered (according to UNESCO Atlas of the World's Languages in Danger)
- Not sufficiently described
- Kullui-speaking area has close and long-standing ties with Russia – The International Roerich Memorial Trust (IRMT) founded in 1993 by Svetoslav Roerich is located in Naggar (Kullu district). Nowadays this place is one of the focuses of the Russian-Indian cultural cooperation.

# Definitely endangered language

- Oral speech mostly, no written codification
- Some Kullu residents consider their native language as being low-status and preventing them from mastering Hindi
- Many parents speak Hindi with their children from childhood, so children do not know their native language

# Under-described language


- Diack A.H. The Kulu Dialect of Hindi: Some notes of its Grammatical Structure, with Specimens of the Songs current amongst People, and Glossary. Lahore, 1896
- Bailey, T.G. The Languages of the northern Himalayas, being studies in the grammar of twenty-six Himalayan dialects. London,1908
- Grierson G.A. Kulu Group // Linguistic Survey of India. Vol.IX. Part IV. Calcutta, 1916.
- Ranganatha, M. R. Survey of Mandeali and Kului in Himachal Pradesh. Census of India, 1971. Language monograph No.7. Delhi, 1981
- Dayanand Sharma (Saraswat). Kuluti-Hindi vyakaran. Ek tulnatmak adhyayan. New Delhi, 2014.
- Maulu Ram Thakur, Pahari bhasha kului ke vishesh sandarbh men. Delhi, 1975.

## Developing a corpus of a minor Indo-Aryan language

- Description of a modern state of minor languages is important considering fast changes under the influence of larger languages and endangerment factors
- Text corpus preserves real language use
- Audio and video recordings can be used in education
- Can be used irrespectively of a linguistic framework of a particular description
- Preserves cultural and linguistic heritage for future generations

# Making a corpus

- Collecting oral texts in different locations of the Kullu district
- Text transcription and translation (into English and Russian, sometimes Hindi) with the help of native speakers of Kullui in ELAN
- Morphological glossing of the texts in Fieldworks Language Explorer
- Making a text corpus with audio recordings
- Corpus studies of grammatical and lexical features of Kullui

# Choice of informants

- Age
- Gender
- Education
- Social status
- Occupation

# Types of texts

*30 texts*

- Religious texts
- Life narratives
- Material culture
- Folklore
- Traditional songs
- Grammatical questionnaires

# Importance of questionnaires

Grammatical questionnaires are included into the corpus. Why?

- All the data on an under-described minor language is valuable
- A few natural texts may lack a lot of grammatical features that can be easily obtained by elicitation
- We consider elicitation and natural texts as complementary sources of the grammatical and lexical data

# Corpus development

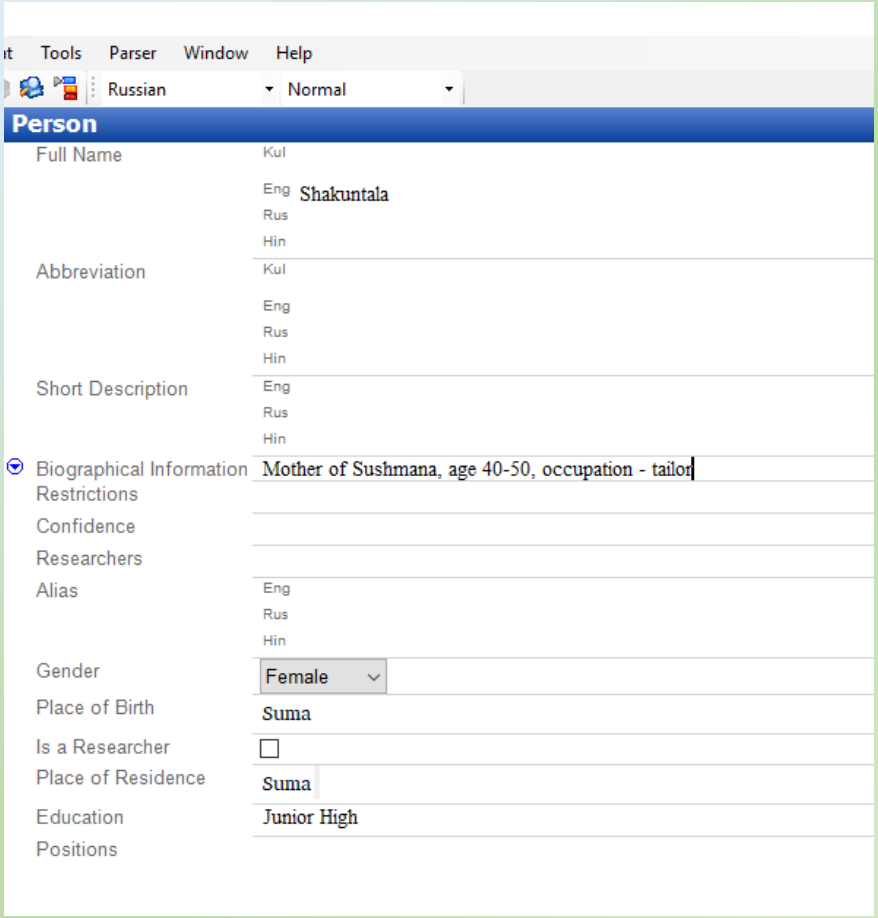
- Metadata
- Transcription and unification
- Morphological analysis and glossing in FLEx
- Internet version of the corpus (next step)

# Metadata

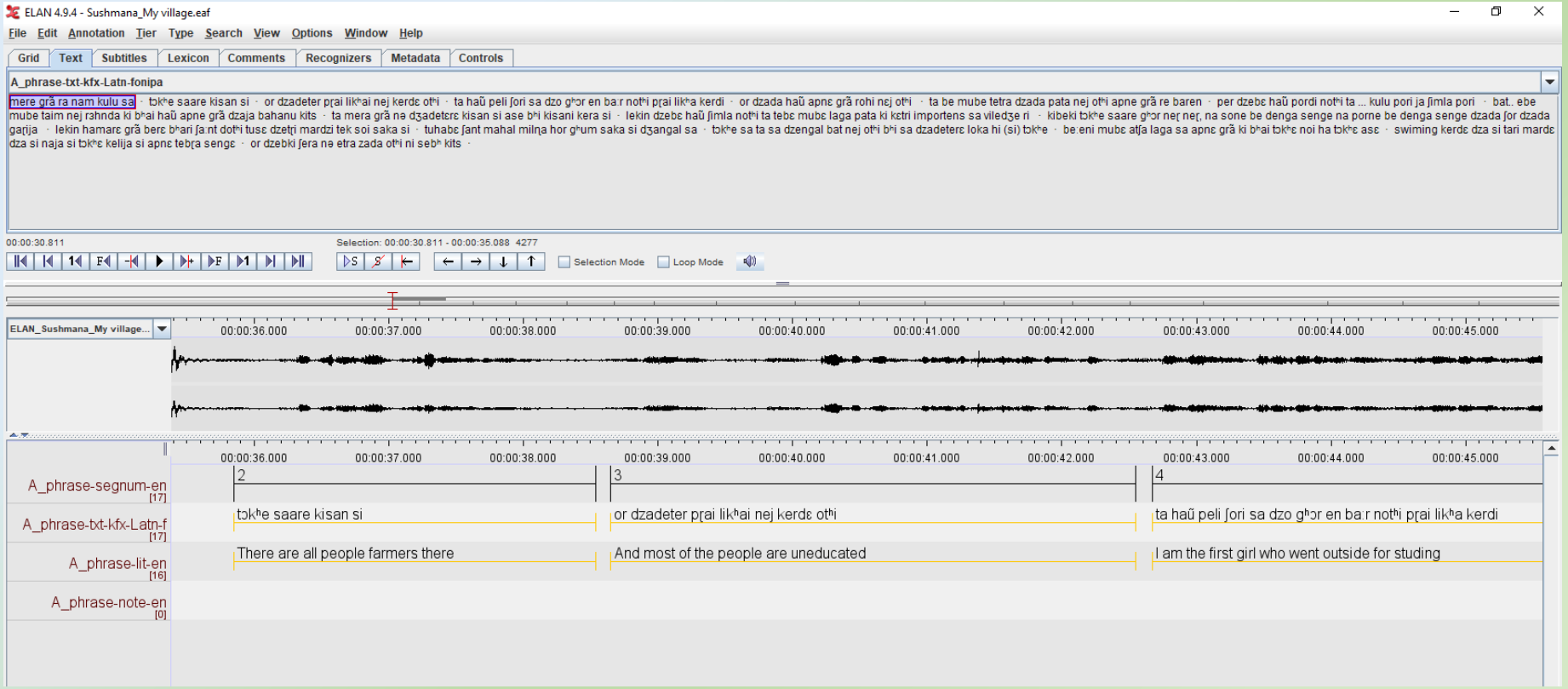
- Place of recording
- Socio-demographic data on the informant: age, gender, education, social position, occupation
- Linguistic history of the informant: place of birth, place of education, linguistic environment, languages of the relatives
- Text genre


505
Хануман Тибба
Hanuman Tibba
Берва
Burwa
Чхатру
Chhatru
Батал
Batal
Башишт
Bashisht
Манали
Manali
Индрасан
Indrasan
Джагатсук
Jagatsukh
Наггар
Naggar
3
Parvati Parvat
Малана
Malana
Маникаран
Manikaran
Tosh
Кхир Ганга
Khir Ganga
Куллу
Kullu
Херанг
Hurang
Мохал
Mohal
Сосан
Sosan
Бантар
Bhuntar
Баяора
Bajaura
3
Шаншар
Shanshar
Машяр
Mashyar
Лихана
Lihana
Banala
Сайнж
Sainj
Aut
Шангар
Shangarh
154
Манди
Mandi
305
3

# Metadata


Tools
Parser
Window
Help
Russian
Normal
Person
Full Name
Kul
Eng
Shakuntala
Rus
Hin
Abbreviation
Kul
Eng
Rus
Hin
Short Description
Eng
Rus
Hin
Biographical Information
Mother of Sushmana, age 40-50, occupation - tailor
Restrictions
Confidence
Researchers
Alias
Eng
Rus
Hin
Gender
Female
Place of Birth
Suma
Is a Researcher
Place of Residence
Suma
Education
Junior High
Positions

# Texts in ELAN



# Morphological analysis in FLEx

- Grammar description and lexicon (by our team)
- Analytism and inflectedness
- Morphology and semantics, GER\IMPF

# Morphological analysis in FLEx

- *hau nei ker-**d**-a* ‘I don’t do’

do-IPFV_PT**NEG_IPFV**-M

- *hau ker-**d**-a au* ‘I came to do’

do-IPFV_PT**PURP**-M

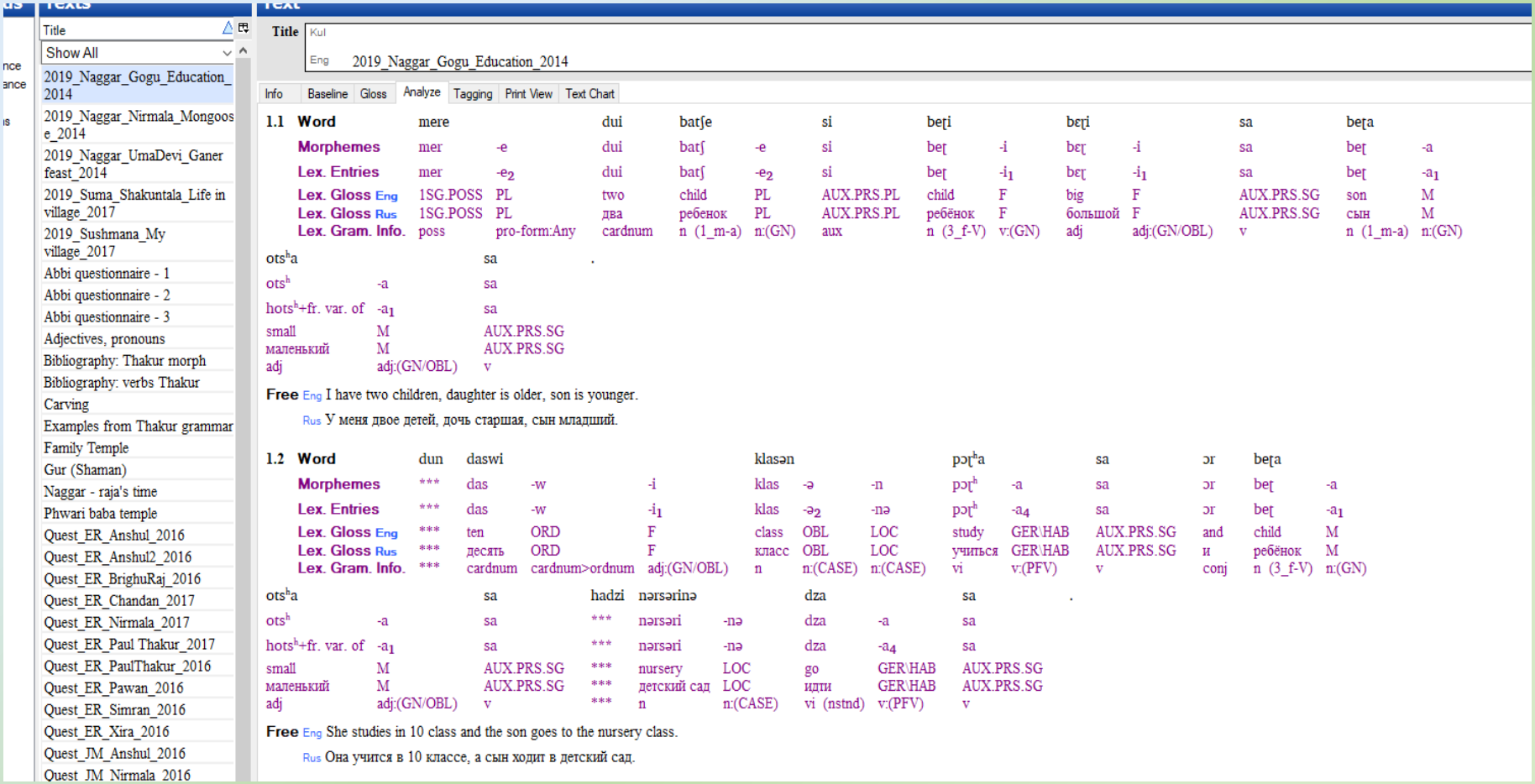
- *heri ker-**d**-a* ‘don’t ever do’

do-IPFV_PT**PROH.FUT**-M

- *hau ker-**d**-a ta* ‘if I do…’

do-IPFV_PT**COND-**M

# Texts in Fieldworks



Texts

Title

Show All

2019_Naggar_Gogu_Education_2014
2019_Naggar_Nirmala_Mongoose_2014
2019_Naggar_UmaDevi_Ganer feast_2014
2019_Suma_Shakuntala_Life in village_2017
2019_Sushmana_My village_2017
Abbi questionnaire - 1
Abbi questionnaire - 2
Abbi questionnaire - 3
Adjectives, pronouns
Bibliography: Thakur morph
Bibliography: verbs Thakur
Carving
Examples from Thakur grammar
Family Temple
Gur (Shaman)
Naggar - raja's time
Phwari baba temple
Quest_ER_Anshul_2016
Quest_ER_Anshul2_2016
Quest_ER_BrighuRaj_2016
Quest_ER_Chandan_2017
Quest_ER_Nirmala_2017
Quest_ER_Paul Thakur_2017
Quest_ER_PaulThakur_2016
Quest_ER_Pawan_2016
Quest_ER_Simran_2016
Quest_ER_Xira_2016
Quest_JM_Anshul_2016
Quest_JM_Nirmala_2016

Text

Title Kul

Eng 2019_Naggar_Gogu_Education_2014

Info | Baseline | Gloss | Analyze | Tagging | Print View | Text Chart

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.1 | **Word** | mere | | | dui | batʃe | | | si | beɽi | | | beɽi | | | sa | beɽa | |
| | **Morphemes** | mer | -e | | dui | batʃ | -e | | si | beɽ | -i | | beɽ | -i | | sa | beɽ | -a |
| | **Lex. Entries** | mer | -e$_2$ | | dui | batʃ | -e$_2$ | | si | beɽ | -i$_1$ | | beɽ | -i$_1$ | | sa | beɽ | -a$_1$ |
| | **Lex. Gloss** Eng | 1SG.POSS | PL | | two | child | PL | | AUX.PRS.PL | child | F | | big | F | | AUX.PRS.SG | son | M |
| | **Lex. Gloss** Rus | 1SG.POSS | PL | | два | ребенок | PL | | AUX.PRS.PL | ребёнок | F | | большой | F | | AUX.PRS.SG | сын | M |
| | **Lex. Gram. Info.** | poss | pro-form:Any | | cardnum | n (1_m-a) | n:(GN) | | aux | n (3_f-V) | v:(GN) | | adj | adj:(GN/OBL) | | v | n (1_m-a) | n:(GN) |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ots$^h$a | | | sa | . | | | | | | | | | | | | | | |
| ots$^h$ | -a | | sa | | | | | | | | | | | | | | | |
| hots$^h$+fr. var. of | -a$_1$ | | sa | | | | | | | | | | | | | | | |
| small | M | | AUX.PRS.SG | | | | | | | | | | | | | | | |
| маленький | M | | AUX.PRS.SG | | | | | | | | | | | | | | | |
| adj | adj:(GN/OBL) | | v | | | | | | | | | | | | | | | |

**Free** Eng I have two children, daughter is older, son is younger.

Rus У меня двое детей, дочь старшая, сын младший.

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.2 | **Word** | dun | daswi | | | | klasən | | | pɔɽ$^h$a | | sa | ɔr | beɽa | | | | | | |
| | **Morphemes** | *** | das | -w | -i | | klas | -ə | -n | pɔɽ$^h$ | -a | sa | ɔr | beɽ | -a | | | | | |
| | **Lex. Entries** | *** | das | -w | -i$_1$ | | klas | -ə$_2$ | -nə | pɔɽ$^h$ | -a$_4$ | sa | ɔr | beɽ | -a$_1$ | | | | | |
| | **Lex. Gloss** Eng | *** | ten | ORD | F | | class | OBL | LOC | study | GER\HAB | AUX.PRS.SG | and | child | M | | | | | |
| | **Lex. Gloss** Rus | *** | десять | ORD | F | | класс | OBL | LOC | учиться | GER\HAB | AUX.PRS.SG | и | ребёнок | M | | | | | |
| | **Lex. Gram. Info.** | *** | cardnum | cardnum>ordnum | adj:(GN/OBL) | | n | n:(CASE) | n:(CASE) | vi | v:(PFV) | v | conj | n (3_f-V) | n:(GN) | | | | | |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ots$^h$a | | | sa | | hadzi | nərsərinə | | | dza | | | sa | . | | | | | | | |
| ots$^h$ | -a | | sa | | *** | nərsəri | -nə | | dza | -a | | sa | | | | | | | | |
| hots$^h$+fr. var. of | -a$_1$ | | sa | | *** | nərsəri | -nə | | dza | -a$_4$ | | sa | | | | | | | | |
| small | M | | AUX.PRS.SG | | *** | nursery | LOC | | go | GER\HAB | | AUX.PRS.SG | | | | | | | | |
| маленький | M | | AUX.PRS.SG | | *** | детский сад | LOC | | идти | GER\HAB | | AUX.PRS.SG | | | | | | | | |
| adj | adj:(GN/OBL) | | v | | *** | n | n:(CASE) | | vi (nstnd) | v:(PFV) | | v | | | | | | | | |

**Free** Eng She studies in 10 class and the son goes to the nursery class.

Rus Она учится в 10 классе, а сын ходит в детский сад.

# Searching in corpus

- Metadata
- Lexemes and grammatical forms of Kullui
- Lexemes and grammatical forms of translation

# Corpus studies

- Dialectal variation
- Language change in grammar and lexicon
- Archaic variants, innovations, borrowings

# Corpus studies

| | Central dialect | Northern dialect | Southern dialect | South-Western dialect |
|---|---|---|---|---|
| s~h variation in some words | + | - | + | + |
| PL of copula | + | - | - | - |
| Perfective participle markers | M -u-da<br>F -i-di<br>PL -ɛ-dɛ | M -u-ra<br>F -i-ri<br>PL -ɛ-rɛ | M -i-ra<br>F -i-ri<br>PL -i-rɛ | M -i-ra<br>F -i-ri<br>PL -i-rɛ |
| Archaic locative in -e | - | - | - | + |
| Instrumental postposition kɛrija | - | + | + | + |

# Central and Northern dialects

| Isogloss | Central dialect | Northern dialect |
|---|---|---|
| Loss of intervocalic l ʁ ɭ | –<br>brali ‘cat’, kʰeliŋa ‘play’, -ŋuaɭa (Agentive noun) | +<br>brai, kʰeiŋa, -ŋuaa |
| Alternation s ~ h in some words | +<br>asɛ ~ ahɛ ‘we’, tusɛ ~ tuhɛ ‘you’, tesa ~ teha ‘she.OBL’, sa ~ ha (copula), sɛɲgɛ ~ hɛɲgɛ (Instumental) | –<br>only asɛ, tusɛ, tesa, sa, sɛɲgɛ |
| PL of copula | +<br>SG sa – PL si | –<br>SG, PL sa |
| Perfective participle markers | M -u-da<br>F -i-di<br>PL -ɛ-dɛ | M -u-ra<br>F -i-ri<br>PL -ɛ-rɛ |
| Present tense in positive form | M -a sa<br>F -a sa<br>PL -a si<br>(gerund + copula) | M -ara ~ -ala<br>F - ɛri ~ -ɛli<br>PL -ɛrɛ ~ -ɛlɛ |
| Interrogative particle -i | –<br>eŋu? ‘May I come in?’ | +<br>eŋu-i? |
| OBL of some pronouns | ei (3SG.M. PROX), tei (3SG.M. DIST), koi ~ koiki ~ kosi ~ kosa ~ koski (Indefinite pronoun) | ie, tie, kosu |

# Example of archaic form

| | | | |
|---|---|---|---|
| tebe | tin$^{h}$-e | | bol-u |
| then | they.DIST-ERG | | say-PFV.M |
| | | | |
| mer-a | **kɔm** | **hu-i** | **go-u** |
| my-M | work | be-i.form | go-PFV.M |
| | | | |
| mu | ebe | lupt | ho-ɳa |
| I.OBL | now | disappeared | be-GERV |
| | | | |
| te | | lupt | hon-d-e lag-e |
| they.DIST | | disappeared | be-PTCP-PL DUR-PFV.PL |

Then they (Krishna – honorative PL) told – my **job has been done** I am going to disappear, and he began to disappear.

*"Standard" expression of perfective meaning will be "**kɔm hu-a**"*
*work be-PFV.M*

# Innovations, loan-translations from Hindi

**Kullui**

- ʃori ro-nd-**i** lag-i
- girl.**F** cry-IMPV_PT-**F** begin-F

A girl began to cry.

**Kullui under the influence of Hindi**

- ʃori ro-nd-**e** lag-i
- girl cry- IMPV_PT-**OBL** begin-F

**Compare Hindi**

- laṛkī ro-n-**e** lagī
- girl cry-INF-**OBL** begin-F

# Conclusion

The corpus of a language belonging to the under-investigated group of Himachali Pahari is an important step towards the documentation and the research of these languages. Audio corpus with translation into Russian and English and search tools for grammar and lexical parameters can be used for the Indological, dialectological, comparative-historical and typological studies.