Team and collaboration

As the Principal Investigator, Henrik Liljegren was responsible for the overall project coordination and implementation, but a large number of people and groups contributed in crucial ways to the project and to this database. The main part of systematic data annotation, processing and analysis took place at Stockholm University. Noa Lange was employed as a research assistant, primarily working with IPA transcription, various types of data processing and ELAN annotation. Initially as part of a project internship, and later as a research assistant, Nina Knobloch fine-tuned data and prepared the output of structural analysis and set up a first version of a web interface. Robert Forkel prepared the final structure and output of the present application.

[Pictures]

HenrikLiljegrenSq.jpg

Henrik Liljegren, PI

NoaLange2Sp.jpg

Noa Lange, Research Assistant

NinaKnoblochSq.jpg

Nina Knobloch, Project Intern/Research Assistant

As part of the data collection process (see below for details on the data set), six collaborative elicitation workshops were arranged under the auspices of three different institutions based in the Hindu Kush region. The PI was present and coordinated the elicitation during all of these sessions.

Forum for Language Initiatives (FLI), Pakistan. FLI is a non-governmental organization based in Islamabad. It was established in 2003 to promote the many indigenous languages in Pakistan’s mountainous northern region. It functions as a resource centre for the local communities, offering training and support in the area of mother-tongue education, language advocacy, literature development and language documentation. Two collaborative elicitation workshops were held in Islamabad in October 2015 and August 2016.

Samar, Afghanistan. Samar is a non-governmental organization. It was established in 2009 and functions as a resource centre for language communities throughout Afghanistan, offering training and support in the area of mother-tongue education, literature development and language documentation. Two collaborative elicitation workshops were held in Kabul and one in Faizabad in April 2017.

Department of Linguistics, University of Kashmir, India. The department hosts a handful of linguistic scholars and some 20 Ph.D. students, many of whom are dedicated to the documentation and description of linguistic diversity and multilingualism in Jammu & Kashmir, a union territory in northern-most India. One collaborative elicitation workshop was held in Srinagar in April/May 2018.

The contributions of those institutions were primarily in providing logistic support and work space, identifying and recruiting language consultants, co-facilitating workshop sessions with the PI, and assisting in e.g. audio or video recording. In some cases, staff members or students were instrumental in translating elicitation material and in digitizing materials (translations, transcriptions or glosses) produced non-digitally by participating language consultants.

In addition to the collaborative workshops, a few individual language elicitation sessions were arranged in Islamabad (Pakistan) in August 2016, in Gilgit in August 2016 (Pakistan), and in May 2018 in Kargil (India).

[Pictures]

SrinagarWorkshop.jpg

Collaborative elicitation workshop in Srinagar (India), April/May 2018: Afreen Nazir (Kashmiri), Mohd Mustafa (Purik)

FaizabadWorkshop.jpg alternatively FaizabadWorkshop-2.jpg alternatively FaizabadWorkshop-3.jpg

Collaborative elicitation workshop in Faizabad (Afghanistan), April 2017

IslamabadWorkshop.jpg

Collaborative elicitation workshop in Islamabad (Pakistan), August 2016: Shahid ur Rehman (Gojri), Raja Hasrat Khan (Hindko), Ghulam Rauf Qurashi (Kundal Shahi), Dr. Uzma Anjum (Pothwari), Luke Rehmat (Kalasha)

GilgitElicitation.jpg

Individual elicitation in Gilgit (Pakistan), August 2016: Aejaz Karim (Wakhi)

Totally 79 native consultants, representing 59 language varieties (see Languages for details), participated in the elicitation workshops/sessions. These individuals offered plenty of linguistic and sociolinguistic insights in interaction with the PI and with other participants, and contributed language data in audio and video recording as well as through writing. To the extent they were able, the consultants made draft (mainly non-IPA) transcriptions, translated portions of text into a language of wider communication, and provided word glosses.

A number of students at the Department of Linguistics, Stockholm University, have carried out thesis work related to the project under the supervision of the PI: Andrew Gomes, Nina Knobloch, Richard Kowalik, Noa Lange, Julia Lautin, Jane Ogawa, Vanessa Quasnik, Hanna (Kjellberg) Rönnqvist and Jacqueline Venetz. Those theses can be accessed in the digital archive of DiVA (Digitala Vetenskapliga Arkivet): su.diva-portal.org

Project design and data collection

With minor exceptions, the dataset collected in each language consisted of seven components: three word lists, a sentence questionnaire, a translated parallel text, context-elicited demonstrative expressions, and a stimulus-based narrative. Apart from the elicitation materials and instructions in English, translations were made available in three major lingua francas: Urdu, Dari and Pashto.

|  |  |  |
| --- | --- | --- |
| Data component | Description | Recorded form |
| 40-list [40list] | Word list including the 40 basic vocabulary items used by ASJP (Wichmann, Holman & Brown 2016) | Written (mostly Arabic-based)+audio |
| Kinship [Kin] | Word list including 95 kin relations, designed by the PI | Written+audio |
| Numerals [Num] | Word list including the cardinal numerals 1-50, 60, 70, 80, 90, 110, 120, 200, 1000 | Written+audio |
| Valency [Val] | Sentence questionnaire representing 87 verb meanings, designed by the by the Leipzig Valency Classes Project (<http://valpal.info/about/project>) | Written+audio |
| North Wind [NW] | Translation of the traditional fable *The North Wind and the Sun*, widely used for illustrating the phonetics of numerous languages (International Phonetic Association 1999) | Written+audio (for a subset: written+audio+video) |
| Demonstratives [Dem] | Expressions used in reference to objects situated at various distances from the speaker. An elicitation kit for this was designed by the PI, largely inspired and guided by Wilkins (1999) | Audio+video (for a subset, the consultants transcribed their own utterances based on the audio recordings) |
| Pear Story [PS] | Natural or semi-natural speech used in retelling the contents of the six-minute “Pear film”; see Chafe (1980) and <http://pearstories.org/> | Audio+video (for most of the varieties, the consultants transcribed their own speech and provided a translation to a lingua franca) |

For the large majority of language varieties (53 of 59), data collection was carried out in the context of a 4-5 days’ workshop, each involving consultants from five or more languages, in which the following procedure was followed: the participants were given a basic introduction to one of the components, e.g. kinship terms, (if applicable) they were given time to prepare themselves for recording (either individually or group-wise) by filling in a word list or questionnaire in whatever style they preferred, they were then invited, each in turn, to a (makeshift) recording studio to be audio or audio-and-video recorded, and after that a considerable amount of time was spent in comparing and discussing particular pieces of data among the participants. All consultant-produced written material was either saved electronically or photo-copied, to aid further processing.

Subsequent to data collection, the material was organized, selectively transcribed (using a broad IPA transcription), analyzed and coded. The data set allowed for classifying each variety according to 80 binary structural features, as presented and detailed under Features, with the cited examples drawn from various components of the set. The elicitation and processing of the three wordlists generated the data under Wordlists. Much of the data remains to be fully annotated and is projected to become available on this site in future installments.

Acknowledgements

This project and the generation of the present dataset would not have been possible without the generous and positive collaboration of the 79 language consultants that took the time to participate in data elicitation sessions in Afghanistan, India and Pakistan. While it is unfortunate, we have for safety reasons (particularly in Afghanistan) decided to let them remain anonymous (apart from people occurring in photos).

We would like to extend thanks to the teams and individuals from the three collaborating institutions in the region. A special thanks to Fakhruddin Akhunzada, Shams Wali Khan, Amir Haider, Naseem Haider and Muhammad Zaman Sagar of the Forum for Language Initiatives, for recruiting consultants and arranging two successful workshops in Islamabad. Thank you so much, Dennis Coyle, Najib Ullah and the rest of the team at the Samar Kabul office, and Katja Müller, Sani Marzban, Adina Muhammad and the rest of the team at the Samar Faizabad office, for recruiting consultants from a number of very remote locations and for providing excellent services and hospitality in connection with the three workshops in Afghanistan. Many thanks to Aadil Amin Kak, Irfan ul Salam and other faculty members at the Department of Linguistics, University of Kashmir, for recruiting consultants and arranging a successful workshop in Srinagar.

Thanks also go to Allauddin Torwali for digitizing materials collected in Pakistan; to Aziz Ali Dad for recruiting consultants and arranging several individual sessions held in Gilgit; and to Irfan ul Salam (again) for joining me (Henrik) on that memorable and breathtaking data collection trip to Ladakh.

Chafe, Wallace L. 1980. *The pear stories: Cognitive, cultural and linguistic aspects of narrative production* (Advances in Discourse Processes, 0896-470X ; 3). Norwood, N.J.: Ablex.

International Phonetic Association. 1999. *Handbook of the International Phonetic Association : A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.

Wichmann, Søren, Eric W Holman & Cecil H Brown. 2016. The ASJP Database (version 17). http://asjp.clld.org/ (19 December, 2017).

Wilkins, David. 1999. The 1999 demonstrative questionnaire: “this” and “that” in comparative perspective. In *Manual for the 1999 Field Season*, 1–24. Nijmegen: Max Planck Institute for Psycholinguistics.