

## Diachrony of linguistic features in South Asian languages

South Asia with its immense diversity of languages belonging to several families and a few language isolates has been considered a classical example of so called linguistic area. A linguistic area has been defined in terms of a number of linguistic features shared by languages of a given region which belong to various stocks. In South Asia these features belong to various language levels: phonology, morphology and syntax (e.g. retroflex consonants, morphological traits such as a lack of verbal prefixes and prepositions, two stems of personal pronouns, conjunctive participles, echo-word constructions, classifiers) and they have been shared by Indo-Aryan, Dravidian, Munda, Tibeto-Burman as well as a few isolates.

Initial research in the middle of the 1950s was a first attempt to build a catalogue of features defining South Asia as a linguistic area which was then further extended in the middle of the 1970s. The beginning of the 21st century brought a major shift of focus to smaller regions of the subcontinent which show a high density of shared features which are not necessarily found on a large scale. However, since the exhaustive set of features has never been proposed, more recent works take a different approach to this issue and abandon the notion of a linguistic area altogether, focusing on the convergence and spread of various features among languages which can but need not necessarily belong to different stocks. This is the starting point for the present project. Most recent areal research on South Asia has focused on large scale analysis of attested features by means of advanced statistical tools. A long-standing desideratum of contact research in South Asia is diachronically focused work on previously established feature-distributional patterns.

The present project thus aims at demonstrating how major linguistic features in early varieties of New Indo-Aryan as well as Dravidian languages for which we have literary sources. These include Nepali, Kashmiri, Braj, Dakkhini, Rajasthani, Gujarati, Marathi, Konkani, Bengali, Maithili, Assamese, Punjabi, Tamil, Kannada and Malayalam. This problem seems to be crucial for better understanding linguistic history of the region.

In the present project we plan to develop diverse corpora of several early varieties of New Indo-Aryan and Dravidian which are attested in many literary genres as well as inscriptions. We plan to trace feature distributional patterns in these texts and then verify whether they have been preserved in modern varieties and how they have developed.

In order to conduct such research not only investigators have to annotate corpora and build corpora of features but they also have to plan field work on lesser known languages and their dialects.

The expected result of the research planned in the project is the reconstruction of the latest linguistic history of the region based on a precise analysis of the development, spread and variation of the linguistic features by means of advanced statistical methods. In total we plan to investigate around 240 features.