

Chapter 2 Research Design

The central design components for this piece of research are its two-stage methodology and the selection of variables and interpretive methods for each stage. A full discussion of the theory and methodologies which guide the interpretation of data is the subject of Chapter 3.

2.1. Two-stage methodology

A two-stage methodology has been adopted in order to address a two-fold challenge in data collection. Firstly, the data should be representative of the linguistic system of individual speakers—their lexicon, phonology, morphology, and so on. Secondly, for the purposes of this study the data should also be as representative as possible of the population of speakers—which runs well into the millions! To collect data that are internally representative of linguistic systems, and externally representative of populations of speakers, is a considerable challenge. While the method adopted here has certain limitations, the data collected have been found adequate to fulfill the purposes of this study.

Stage 1: more linguistic data, less informants

The first stage began with the collection of large amounts of linguistic data at 8 collection sites across KRNB. The data collected were a list of over 1300 lexical and morphological items. These are presented in large part in Appendix A in the form of a comparative wordlist. The wordlists have been transcribed phonemically, with key allophones also distinguished. The phonological analysis which guides the transcription can be gathered from the descriptive sections of Chapter 4. Attention has been given to the phonological analyses of the previous studies listed in 1.7, as well as fresh analysis undertaken as part of this study. In this way, the study adheres to Southworth's criterion that "[w]here possible, the descriptive material on which a comparative treatment is based should be phonemically accurate, and should represent the actual usage of speakers of a particular dialect." (Southworth 1958: 56).

The focus of comparative reconstruction in this project is on phonological forms of lexical items as well as morphological form-function pairings. Detailed

reconstruction of meaning changes typically occurs subsequent to the reconstruction of forms, and has not been undertaken as part of this study. Consequently, the data in the Appendices should be viewed as a precise representation of the forms of etyma, with the English glosses serving only as approximations for the meaning of each item in each lect. Despite being approximations, the meanings given are judged to be accurate enough to justify cognacy.

The lexical items collected were selected so as to represent different semantic domains and syntactic classes of the lexicon:

- Non-animate nature;
- Animals;
- Human body;
- Human food;
- Human experiences/emotions;
- Human products;
- Human relationships & kinship terms;
- Human jobs;
- Village household terms;
- Parts of the banana tree (culturally important);
- Religious terms;
- Diseases, disabilities, illnesses;
- Postpositions, Conjunctions;
- Numbers;
- Days of the week, months of the year;
- Adjectives;
- Verbs, including causatives.

The morphological items collected include the following sets:

- Nominal inflections of case, noun class and number;
- Personal pronouns;
- Pronominals of quantity (e.g. ‘this many’) & quality (e.g. ‘like this’);
- Temporal and locational pronominals (eg. ‘here’, ‘where’, etc.);
- Verbal inflections for various categories of tense-aspect and subject agreement.

Those who assisted with this first stage of data collection were people who had been raised at one of the 8 collection sites, whose parents had spoken the local lect as their mother tongue, and who had sufficient interest in their local lect to sustain data elicitation over several days. Biodata for the informants at each site of this first stage of data collection are given in Appendix B.

The location of each of the test sites was selected firstly on a geographical basis, with an approximately equal distance between each site. Secondly, the selection of sites was also determined by the precise location of available, willing and appropriate informants. The locations of the selected sites are marked in Figure 2-1; the names of the locations are, from west to east:

- Rangeli, located in the south-east of Morang district of Nepal. Speakers here use the terms ‘Rajbanshi’ and ‘Tajpuria’ to refer to their mother tongue, depending on the caste of the speaker. ‘Rajbanshi’ is the name officially recognised by His Majesty’s Government of Nepal.¹
- Kishanganj, located in the north-east of Bihar state, India. Speakers here use the term ‘Surjapuri’ to refer to their mother tongue.
- Mahayespur, located in the east of Jhapa district of Nepal. The same language names apply as for Rangeli. The Mahayespur lect bears many similarities with the lects used over the border in the south-west corner of Darjeeling district in India (Toulmin 2002).

¹ Renamed by the Parliament of Nepal as the Government of Nepal, on Thursday 18th May, 2006.

- Thakurgaon, located in Bangladesh, north of the town of Dinajpur, within the district of Thakurgaon. The precise area of data collection was near to Lohaghara. Speakers in this area refer to their mother tongue as ‘Deshi bhasha’, ‘Dhekri’ or ‘Dhekia’ (cf. 1.6).
- Shalkumar, located in the centre of Jalpaiguri district, India. Speakers of this area refer to their mother tongue variably as ‘Kamta(puri)’, ‘Rajbanshi’ or ‘Deshi bhasha’ as described in 1.6.
- Rangpur, the headquarters of the district. During this first stage of research, the data were collected with speakers at several sites outside of the town area, cf. Appendix B. Speakers of this area refer to their mother tongue as either ‘Bahe’, ‘Rangpuri’, ‘Deshi bhasha’ or its synonym ‘Anchalit bhasha’ meaning “the local language”.
- Bhatibari and Tufanganj. The first town is located in the extreme south-east of Jalpaiguri district on the border of Cooch Behar district. The second town is found south of Bhatibari within the confines of Cooch Behar district and near the border with Bangladesh.
- Bongaigaon, the headquarters of a district in Assam state. The data were collected with speakers residing in old Bongaigaon town.



Figure 2-1. Location of sites for the first stage of data collection

The data collected under stage 1 were analysed using the Comparative Method of historical linguistics. Methodology for phonological reconstruction is discussed in Chapter 3. For methodology of morphological reconstruction, see Koch (1996). The results of the phonological and morphological reconstruction are given in Chapters 4-6. These results in turn guided the selection of linguistic variables for Stage 2 of data collection and the sociohistorical reconstruction that results from these data in Chapter 7.

Stage 2: less linguistic data, more informants

The second stage of the research design involved the selection of phonological and morphological variables identified by the comparative reconstruction based on Stage 1 data. Under Stage 2, a smaller amount of data targeting these linguistic variables was collected at a larger number of sites from a greater number of speakers. The dependent linguistic variables were selected based on the following criteria:

- Type of change: selected so as to include phonological, nominal morphological and verbal morphological lexical innovations.
- Diagnostic value of the change: the linguistic variables tested were those reconstructed as diagnostic of propagation events (cf. 3.4.1) and thus useful for the sociohistorical reconstruction of linguistic history in Chapter 7.

Informants for this second stage were sampled so as to represent the population of speakers in terms of the following independent variables:

- Geography: the collection sites were located much closer together, with a distance of 25-30 kilometres separating each site. The geographical locations are shown in Figure 2-2 each given a unique numerical label. The names of the village or town at which data were collected during this stage are given in Appendix D.
- In the case of site 24, as well as some other sites, Muslims and Hindus were found to be living in close proximity. Several speakers were interviewed from each of the two socio-religious communities, though not all the data have been analysed as yet.

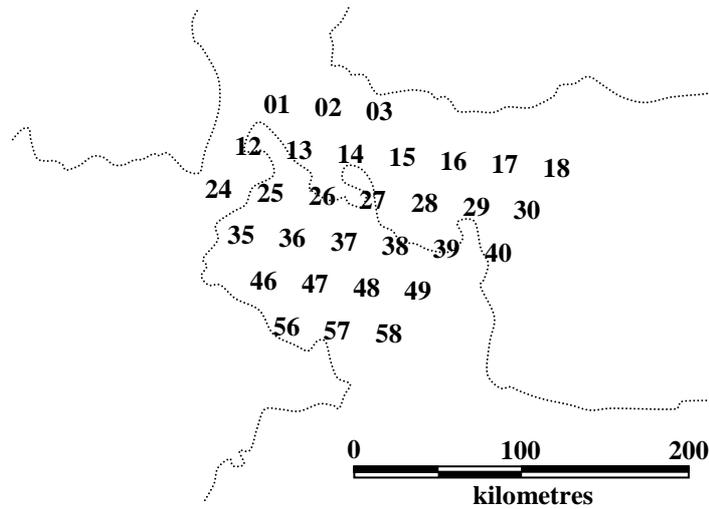


Figure 1-2. Geographical location of the test sites for Stage 2 data collection

The informants at each site were controlled with respect to the following variables:

- Locally based: each informant was presently living in the village, and was born and raised within 5 kilometres;
- Continuing association with local rural life: each informant identified themselves as *sthaniyo manshi* ‘one of the locals’;
- Language spoken in their home: each informant identified the language that they speak in their home under one of the naming schemes described in Chapter 1 for the KRNB lects.
- Education level: informants for the collection of lexical and nominal morphological data were either uneducated or minimally educated (up to around 5 years). For collection of verbal morphological paradigms, more educated speakers were required who could see the point of the abstract activity of keeping one grammatical category constant (tense-aspect) while changing another category (grammatical subject). Ie. ‘I do, you do, he does, etc.’
- Mobility/stability: each informant had not lived away from the village for more than 5 years.
- Age: wherever possible, each informant was between 30 and 55 years old.

Data collection for this second stage was carried out entirely monolingually, and making use of pictures for collection of lexical data. This contrasted with the Stage 1

data collection which involved a much greater dependency on English for elicitation purposes.

The following variables were intentionally left uncontrolled during Stage 2 data collection:

- Sex: language variation along sex-based lines is not part of this study as such variation has not been found significant enough to be of use for reconstructing linguistic history. An initial pilot test of 10 men and 10 women near the road from Siliguri to Jalpaiguri town showed no significant variation between the sexes. In this regard the results from the pilot concur with those of Toulmin (2002). In that study interviews on perception of lectal variation were carried out with 110 KRNB speakers at 11 villages across northern West Bengal, and Jhapa district of Nepal. In none of those interviews did participants suggest that sex was a significant factor in lectal variation. The responses instead focussed on geographical variation and social variation between religious groups—the two categories selected as independent variables for the present study. During the present study, little variation has been found along sex-based lines. The exception is at site 17 (near Gosaigaon in Assam). At that site, progressive raising of *a was present for female speakers and absent for male speakers (see Appendix D).
- Attitude towards mother tongue: With language/dialect status such a politicised issue in West Bengal and Assam, the sensitive nature of such questions was considered too threatening to the informants, with potential to cause them unnecessary stress and put the reliability of the results in doubt. While attitudinal factors have been shown to have a significant bearing on language variation and change (Marshall 2004), the goal of this study is not the description of language variation in progress but rather the reconstruction of linguistic history. The factors that are most relevant to this study therefore are not the attitudes of present-day speakers, but rather the attitudinal conditions that were in effect during the time when the linguistic innovation—now an embedded stable feature—was a change in progress. Thus the omission of present-day attitudinal factors as independent or controlled variables in this study does not threaten the reliability of the results.

In total, interviews were carried out at 30 different sites during Stage 2, with 4 speakers interviewed at almost all sites. The schedule of biodata questions and the concepts elicited in these interviews are given in Appendix C. The results of this dialectological research are given in full in Appendix D, and included in the historical argument in Chapter 7 in accordance with the methodology developed in the next Chapter.