From linguistic to sociolinguistic reconstruction: the Kamta historical subgroup of Indo-Aryan

Pacific Linguistics 604

Pacific Linguistics is a publisher specialising in grammars and linguistic descriptions, dictionaries and other materials on languages of the Pacific, Taiwan, the Philippines, Indonesia, East Timor, southeast and south Asia, and Australia.

Pacific Linguistics, established in 1963 through an initial grant from the Hunter Douglas Fund, is associated with the Research School of Pacific and Asian Studies at The Australian National University. The authors and editors of Pacific Linguistics publications are drawn from a wide range of institutions around the world. Publications are refereed by scholars with relevant expertise, who are usually not members of the editorial board.

From linguistic to sociolinguistic reconstruction: the Kamta historical subgroup of Indo-Aryan

Matthew Toulmin

Pacific Linguistics

Centre for Research on Language Change

Pacific Linguistics
Research School of Pacific and Asian Studies
The Australian National University

Published by Pacific Linguistics
Research School of Pacific and Asian Studies
The Australian National University
Canberra ACT 0200
Australia

First published 2009

National Library of Australia Cataloguing-in-Publication entry:

Author:	Toulmin, Matthew
Title:	From linguistic to sociolinguistic reconstruction: the Kamta historical subgroup of Indo-Aryan.
ISBN:	9780858836044 (pbk.)
Notes:	Bibliography.
Subjects:	Rajbangsi language — History. Reconstruction (Linguistics).
Other Authors/ Contributors:	The Australian National University. Research School of Pacific and Asian Studies. Pacific Linguistics.
Dewey Number:	491.44

Copyedited by Melissa Crowther
Typeset by Jeanette Coombes
Cover design by Julie Manley
Printed and bound by Addcolour Digital Pty Ltd, Fyshwick, Canberra

Table of contents

Acknowledgments

The completion of this book was made possible by the patient assistance of many speakers, academics, and my friends and family. Their contributions have added to its strengths in different ways, while in no way absolving me from responsibility for its weaknesses.

Firstly, my thanks go to those who supervised my PhD research, on which this book is based: Dr Harold Koch, Prof. Malcolm Ross, Dr Bethwyn Evans and Dr Luise Hercus of The Australian National University, and Dr Ichhimuddin Sarkar of the University of North Bengal. Their comments on earlier drafts were extremely helpful, and their ideas and suggestions are scattered throughout the study. Other academics who advised me informally over the course of the research project were: Professors Nirmal Das and Pabitra Sarkar of Kolkata, Dr Shafiqul Islam of Rajshahi, Professor Maniruzzaman of Chittagong, Professors Bill Labov and Frank Southworth of Philadelphia, and Dr S.N. Barman of Bhubaneswar. Dr Brendan Whyte kindly gave permission for some of his maps to be reproduced in Chapter 7.

I am grateful to the Faculty of Arts, The Australian National University, which paid for much of the costs of field work, and to the Department of History, University of North Bengal for accepting me as an affiliated researcher while I was living in North Bengal.

I wish to thank all the KRDS speakers who spent hours talking with me, who housed and fed me, who sang to me, and shared their language and lives. I can only name a small number of them here, and many others also deserve credit. In West Bengal, I am especially grateful to Mr Dharma Narayan Barma, Mr Sujan Barman, and Mr Khagendra Nath Ray and their families. Their insight into their mother tongue is exceptional. My thanks also to Mojen Barman, my language teacher and friend. In Nepal, I wish to thank the family of Mr Phul Singh Rajbanshi, especially his son Manohar and in-laws from Darbesha. In Bongaigaon (Assam), I was greatly helped by Mr Rajendra Narayan Deb. In Rangpur, I was aided by Mr Matiur Rahman Basania, Mrs Akhtar Banu, Sayeed, and the president and membership of Abijatrik (a local literature society). Also from Bangladesh, I wish to thank the families of Mr Lalon and Mrs Ripa Noor, and finally Mr Gulam Kibria who introduced me to Bangladesh. If God wills, I hope to meet you all again without too much delay.

I could not have completed this book without the personal support of so many, most of all my wife. Donna, you travelled around the world with me and we experienced many difficulties and adventures together. You are my life companion and closest friend.

Glory to God, and to his Son, Jesus Christ, the author of history.

Linguistic abbreviations and conventions used

Glossing rules and abbreviations follow the Leipzig conventions:

http://www.eva.mpg.de/lingua/resources/glossing-rules.php

1, 2, 3	first, second, third person
A	agent-like argument of canonical transitive verb
ABL	ablative
ACC	accusative
ADJ	adjective
ADV	adverb(ial)
AGR	agreement
AN	animate[1]
ANP	anaphoric[1]
AUX	auxiliary
BEN	benefactive
CLF	classifier
CMP	comparative[1]
DAT	dative (in this study, generally includes Dative-Accusative)
DEF	definite
DEM	demonstrative
DIST	distal
FUT	future
GEN	genitive
H	high (in terms of honour), contrasts with L (low)[1]
IMP	imperative
INAN	inanimate[1]
INDF	indefinite
INF	infinitive
INS	instrumental

[1] These abbreviations are not part of the Leipzig conventions.

INT	interrogative[2]
INTR	intransitive
IPFV	imperfective
L	low (in terms of honour), contrasts with H (high)
LOC	locative
MI	morphological innovation[2]
NOM	nominative
NUM	numeral[2]
NT	neutral (in terms of honour), contrasts with high and low[2]
OBL	oblique
P	patient-like argument of canonical transitive verb
PDL	propagation-defined lect (see §3.2.1)[2]
PE	propagation event (see §3.2.1)[2]
PFV	perfective
PI	phonological innovation[2]
PL	plural
PN	propagation network (see §3.2.1)[2]
PROX	proximal/proximate
PRS	present
PST	past
PTCP	participle
r̥	syllabic ‘r’ as transcribed in Indic studies (cf. §3.3.1)
REL	relative
S	single argument of canonical intransitive verb
SBJ	subject
SC	speech community[2]
SCE	speech community event[2]
SG	singular
TR	transitive
*(*x*)	the presence of *x* in the reconstructed datum is ambiguous
*(*x; y*)	*x* **or** *y* was present for the reconstructed stage
*[*x*]	the datum is reconstructible in two forms, one with *x* and one without *x.*
*[*x; y*]	the datum is reconstructible in two forms, one with *x* and one with *y.*
{X,…}	innovation pertains to lects *X*,…
‘x’; ‘y’	Synchronic variation between form ‘x’ and ‘y’.
(x)	In synchronic form, indicates an optional element ‘x’.
[x]	In synchronic form, indicates a phonlogically conditioned element ‘x’.

[2] These abbreviations are not part of the Leipzig conventions.

Abbreviations for languages, locations and social groups

*	Reconstructed form
BH	Bhatibari, West Bengal, India (site of data collection during Stage 1)
Bhj	Bhojpuri lect
BN	Bongaigaon, Assam, India (site of data collection during Stage 1)
CB	Cooch Behar
c‘X’	Central ‘X’ (e.g. cKRDS)
cMg	Central Magadhan subgroup of lects (along with wMg termed the ‘Bihari’ lects, including Maithili)
ext-‘X’	Extended ‘X’ (e.g. ext-cKRDS)
e‘X’	Eastern ‘X’ (e.g. eMg)
eMg	Eastern Magadhan subgroup of lects (Oriya, Bangla, Asamiya, KRDS, Mal Paharia, etc.)
H	Hindu (socioreligious group)
Hn	Hindi (language)
IA	Indo-Aryan
JL	Jalpaiguri
Kmr	Kamrupi
KRDS	Kamta, Rajbanshi, Deshi and Surjapuri subgroup of Indo-Aryan
KS	Kishanganj, Bihar, India (site of data collection during Stage 1)
M	Muslim (socioreligious group)
Mg	Magadhan subgroup of lects, further divided into eMg, cMg and wMg. (Chatterji 1926)
MH	Mahayespur, Jhapa, Nepal (site of data collection during Stage 1)
MIA	Middle Indo-Aryan
Mth	Maithili
nw‘X’	North western ‘X’ (e.g. nwKRDS)
NIA	New Indo-Aryan
Npl	Nepal (nation) and Nepali (lect)
OIA	Old Indo-Aryan
p	Reconstructed ancestral (or proto) language stage (e.g. pKmt)

pKmr	Proto Kamrupa
pKmt	Proto Kamta. Reconstructed stage in linguistic history ancestral to modern KRDS lects, defined by linguistic propagation events.
Rjb	Rajbanshi
RL	Rangeli, Morang, Nepal (site of data collection during Stage 1)
RP	Rangpur, Bangladesh (site of data collection during Stage 1)
SC	speech community
SCA	Standard Colloquial Asamiya
SCB	Standard Colloquial Bangla
SCH	Standard Colloquial Hindi
SCO	Standard Colloquial Oriya
SH	Shalkumar, West Bengal, India (site of data collection during Stage 1)
Sjp	Surjapuri
TH	Thakurgaon, Bangladesh (site of data collection during Stage 1)
w‘X’	Western ‘X’ (e.g. wKRDS)
wMg	Western Magadhan subgroup of lects (along with cMg, also termed the ‘Bihari’ lects, including Bhojpuri)

1 *Introduction*

The purpose of this book is twofold: to present an innovative sociolinguistic method for reconstructing language history, and to apply this method to the history of an Indo-Aryan subgroup. The subgroup at issue is made up of several lects,[1] spoken in the plains to the south and south-west of Bhutan—an area now divided between India, Bangladesh and Nepal. The findings of this study justify reconstructing a 'proto Kamta' stage of linguistic history, which is the point of common origin for these lects, and defines them as a subgroup within Indo-Aryan. (See §1.6 for a summary of the key historical findings of this study.)

1.1 Issues in reconstructing Indo-Aryan linguistic history

The linguistic history of the Kamta subgroup has been complex, with periods of phylogenetic splitting (a lect splitting into multiple lects), as well as periods of phylogenetic reticulation (several lects undergoing common innovations). Complexity of this sort comes standard in Indo-Aryan (as in dialect continua generally); this is due to the presence of non-discrete boundaries between speech communities. These non-discrete boundaries pose a major challenge to the reconstruction of linguistic history: if two related languages may undergo common innovations *both before and after* they have differentiated from each other, *how is it possible for us to reconstruct the sequencing in which the changes occurred?*

The historical linguistic literature dealing with Indo-Aryan may be divided into three categories based on how each study deals with this problem of chronological ambiguity:

1. *text-based* studies, which reconstruct the chronology of linguistic changes based on the presence or absence of these changes in an historical corpus of written literature;
2. *tree-based* studies, which assume linguistic phylogeny to be primarily 'tree'-shaped, and so reconstruct periods of common innovation followed by language splitting and differentiation;
3. *unsequenced* studies, which acknowledge the ambiguity involved in reconstruction, and so attempt no chronological account of changes.

[1] The term lect is used throughout this study as a synonym for 'speech variety' and refers to 'a language' or 'a dialect' without discriminating between the two. See §1.3 on the sociopolitical distinction between 'a language' and 'a dialect.'

Indo-Aryan has been the subject of some major historical linguistic studies. These have employed, by and large, the text-based approach, guided by the classical methods of philology and etymology (e.g. Bloch 1920; Chatterji 1926). The text-based approach is not without its problems. Most notably, this approach can only be applied to a small number of the Indo-Aryan lects; for only a small number possess a substantial corpus of historical writings.

Even for those languages with an historical corpus, the method remains fraught with problems. Notable amongst these are:

(a) the difficulty of drawing conclusions from ancient writings about the vernaculars of the time when the writings are often intentionally archaic and artificially distanced from spoken norms;[2]

(b) the difficulty of adjudicating between rival claims to linguistic 'ownership' of an ancient text when the linguistic histories are characterised by interconnectedness rather than discrete divisions.

The case of the *Caryapadas* is a famous example of the latter problem, and is particularly pertinent to this study. The *Caryapadas* are an early New Indo-Aryan (NIA) collection of Buddhist mystic songs. They have variously been claimed to represent 'Old Bengali,' 'Old Oriya,' 'Old Maithili,' 'Old Asamiya,' and 'Old Kamta'—invariably by scholars belonging to the language group in question.

Dependence on ancient texts is problematic, but the main alternative to this approach—the Comparative Method of historical reconstruction—is not necessarily any better placed to deal with the ambiguities of dialect continuum history. The Comparative Method has been only sparingly used in historical Indo-Aryan studies (e.g. Southworth 1958; Pattanayak 1966; Maniruzzaman 1977), and has been found wanting in important respects. The limitations and problems connected with this method are discussed in detail in §3.3. To summarise what is argued there: the Comparative Method can reconstruct bare linguistic changes, but all too often the *sequencing* of these changes cannot be established by that method alone. The Comparativist is then faced with the three options outlined above: (1) go to historical texts to establish sequencing (e.g. Southworth 1958); (2) assume a tree-shaped phylogeny (e.g. Pattanayak 1966); or (3) leave the changes unsequenced (e.g. Maniruzzaman 1977). The problems and limitations facing the first option have already been given. The second option is not reliable for historical reconstruction given the potential for phylogenetic reticulation (that is, common innovation occurring after periods of differentiation, cf. §3.2). The third option (which abandons the goal of sequencing changes) concludes with a diagram of overlapping isogloss boundaries, but no coherent historical account (cf. Maniruzzaman 1977). Though conceptually valid, this approach is, from an historical perspective, less than optimal.

The proposal put forward in this book is that the methodological strengths of philology, etymology, the Comparative Method, and dialect geography, be synthesised within a framework provided by the sociohistorical theory of language change. This methodological framework can be applied to lects with unrecorded and recorded histories alike; it can (in

---

[2] Katre comments: 'Like OIA which continued to flourish as a language of literature when MIA was the general channel of communication among the people, MIA in its turn appears to have been used for literary purposes long after it ceased to be current as a common medium of communication. This explains the highly artificial character of MIA literature and its production since 1000 AD, particularly in its latest phase of Apabhraṁśa' (Katre 1968:3).

many instances) disambiguate the sequencing of changes reconstructed by the Comparative Method; and it can reconstruct the complex interconnections between linguistic histories without undue reductionism.

1.2 The speakers and their lects

The lects examined in this book were treated as a unit in the *Linguistic Survey of India* under the title 'Rajbanshi,' which Grierson classified as a 'dialect of Bengali' (Grierson 1903–28). However, Grierson's classification and nomenclature are either unacceptable or unknown to the vast majority of speakers today (cf. §1.2.2). Throughout this book, the lects are referred to as KRDS, which stands for *Kamta, Rajbanshi, Deshi, Surjapuri.* This complex nomenclature reflects the variety of names used by speakers of this set of related lects (cf. §1.2.2).

The KRDS lects are classified in the Ethnologue as *Indo-European, Indo-Iranian, Indo-Aryan, Eastern zone, Bengali-Assamese* (Gordon 2005). This study shows that the final subgrouping of 'Bengali Assamese' is as yet without a firm historical linguistic foundation (cf. §7.3.3).

The geographical extent of the KRDS linguistic cluster includes Indo-Aryan lects spoken in several districts of Nepal, India and Bangladesh, see Figure 1.1

Figure 1.1: The region where KRDS lects are spoken, with international boundaries and district names marked

The population of speakers includes many distinct social, religious and ethnic identities, including (in alphabetical order): *Deshi* ('local') *Muslims* (who also identify themselves as *Bangalis*), *Gangais, Meches, Rajbanshis*, and *Tajpurias.*

The adjective *Deshi* 'local' is important enough to this study to warrant a brief discussion. This term is a marker of in-group status, and distinguishes the indigenous mainstream population from the *Adivasi* 'tribal, aboriginal' (including Bodos and Santalis), and the

Bhattia. The *Adivasi* are indigenous, but not mainstream; the *Bhattia* are migrants from the Bangali South and thus not indigenous to the region. The *Bhattia* identify themselves straightforwardly as *Bangalis* and speak lects much more similar to Standard Colloquial Bangla. Table 1.1 summarises the political areas in which each socioreligious group is found.

Table 1.1: Distribution of socioreligious groups in the KRDS region

Socioreligious group	Geographical area				
	Nepal	Bihar	West Bengal	Bangla-desh	Assam
Rajbanshi	✓	✓	✓	✓	✓
Tajpuria	✓	✓	×	×	×
Gangai	✓	✓	×	×	×
Deshi Muslim	✓	✓	✓	✓	✓
Poliya Hindus	×	×	✓	✓	×

The social diversity of the KRDS speech community finds its parallel in the diversity of names used by speakers for their mother tongue. The issue of naming is discussed in §1.2.2, after an introduction to some relevant sociohistorical dynamics.

1.2.1 Overview of key social dynamics

Language history across north India is time and again linked with *religious conversion*. The importance of Sanskrit to Hinduism has provided a general momentum for non-Aryans who embrace Aryan religion to also embrace (Indo-)Aryan language. In the case of the KRDS lects we must not only consider mass conversion to Hinduism, but also the more recent mass conversion to Islam of at least half the speaker population. The latter conversion process, while leading to a small increase in the use of Persian and Arabic origin vocabulary, has had nothing like the linguistic impact of the earlier conversion to Hinduism.

Islamic conversion has, nonetheless, had a major *social* and *sociolinguistic* impact. The expansion of Islam into the KRDS speaking area came from the Bangali (Bengali) south, and resulted in an *increased identification by converts with that Muslim (and Bangali) south*. Consequently, though the Muslims of Rangpur and the Hindus of Cooch Behar (Koch Bihar) speak highly similar Indo-Aryan lects, they are highly dissimilar in how they understand their social identity, and the identity of their mother tongue. Muslims in Rangpur consistently identify themselves as Bangalis, and conceive of their mother tongue as included within 'the Bangla language'. It is no doubt also relevant that these same speakers joined the rest of their nation in fighting the war of independence against Pakistan. A key demand of that uprising was the recognition of *bangla bhasha* 'the Bangla language' alongside Urdu as a language of administration. The KRDS speaking Hindus, by and large, do not share the same sense of commonality with the Bangali south as their Muslim counterparts. Most of the Hindu speakers identify themselves as 'Rajbanshis,' which brings us to the second sociohistorical dynamic: autonomy (and its loss).

The term *Rajbanshi* is derived from Sanskrit and means 'royal race', or 'descendants of the King'. The term refers back to the autonomous kingdom established in the 16th century under the Koch kings, of which more will be said in Chapter 7. Under the reign of the

Koch kings, even up to AD 1950, this kingdom maintained its general autonomy from the Bangali south and the Asamiya east, though the kingdom's size was gradually reduced over the centuries. Looking further back in history than the Koch kings, the KRDS area was even then autonomous and distinct from the kingdom of Gauda (which later became Bengal). From the 13th to 16th centuries, the kingdom was referred to as *Kāmatā*, and ruins of the old capital of *Kāmatāpur* may be visited today just south of Cooch Behar town.

The Hindus of today's north Bengal are keenly aware of their generally autonomous social history. As a result, there is an increasing use of the term 'Kamtapuri' to denote both a person of local origin—a 'son of the soil'—and the language of local origin, KRDS, as spoken by Hindus and Muslims alike. Understandably, the term 'Kamta' as a language name is politically controversial, implying as it does linguistic autonomy, rather than heteronomy under 'the Bangla language'.

In sum, both the historical situation, and the present situation of the KRDS lects, involves *conflict over sociolinguistic ideas.* In general the conflict is between KRDS speakers who believe in their own sociolinguistic *autonomy*, and speakers of closely related lects (Bangla and Asamiya) who believe in the *heteronomy* of KRDS—'your mother tongue is a dialect of our language'. As stated above, Muslim speakers identify with the Muslim south and its Bangali identity to a greater degree than the Rajbanshi Hindus—with the result that the language conflict is today restricted to the Indian side of the border. Unlike the monsoon floods which sweep down from West Bengal into northern Bangladesh, the sociolinguistic conflict on the Indian side is yet to produce any run-off into Bangladesh.

These three social dynamics—conversion, autonomy (and its loss), and conflict—account for the social and political sensitivity of the subject matter for this study. They also help explain the difficulty in naming the lects. The terms 'Rajbanshi' and 'Kamta' have wide circulation in India and Nepal, but not in Bangladesh. To label all the lects as 'deshi Bangla' would match the ideology and sentiments on the Bangladeshi side, but widely offend on the Indian and Nepali sides. *From a synchronic sociolinguistic perspective, the lects are not well suited to a unitary, overarching 'language name'. From a diachronic perspective, however, they will be shown in this study to share a common historical origin*—termed proto Kamta for reasons outlined in §1.6.

1.2.2 Naming the lects

The question of how to name the contemporary lects is not of great importance to an historical study such as this, because the linguistic realities and histories are the same whatever name we give them. However, as we will be constantly using some language name or other, an overview is offered below of the different terms used by speakers, with some discussion of their historical and contemporary connotations.

Bahe: This name is favoured in Rangpur and adjoining districts. It is derived from a local word used in male greetings, with a function similar to 'mate' in the colloquial Australian greeting 'G'day mate'. Grierson (1903–28) claimed this term referred specifically to the Darjeeling 'sub-dialect' of 'Rajbanshi' but this is a false assertion. 'Bahe' is a term used right across North Bengal, and most prominently in Rangpur, not Darjeeling. Clark (1969) notes that south Bangalis use this term to mock the language of the Rangpuris. Forty years on, it continues to be a term used by outsiders for derision of local Rangpuris. The Rangpuris, however, have embraced the term *Bahe* with pride as a mark of their sociolinguistic distinctiveness.

Deshi bhasha: This term for the language is favoured in all areas, especially amongst people who wish to be non-committal on the political controversy. The term means 'the language of the *desh* (nation, region, locality)'. Its reference is thus too generic and of little use for distinguishing KRDS from other lects—because every language is, in its own region, the 'deshi bhasha'. In Orissa, for example, there is an Indo-Aryan variety spoken by Adivasis ('aboriginals, tribals'); it is termed in a similar manner as 'Desiya Oriya' (cf. Gordon 2005).

Dhekia, Dhekri: These terms are favoured in the north Dinajpur area of Bangladesh—the former variant among Hindus, and the latter among Muslims. Interestingly, U. Goswami (1970) and Grierson (1903–28) mention a very similar name *Dhekeri* for the western Asamiya lect, Kamrupi. Around Kamrup the meaning of this term is disdainful, but the same does not hold in Dinajpur.[3]

Kam(a)ta, or **Kamtapuri**: These terms are favoured in West Bengal by two groups of people: (1) those in favour of the establishment of political autonomy in north Bengal; or (2) those who insist on a language name that is not caste-based. Barma (1991, 2000) is a key advocate for a non-caste-based name for the language. Barma and others have criticised the term Rajbanshi as being too caste-centric, and excluding Muslim speakers. While this argument has some force, the term Kamta(puri) has not been accepted by the West Bengal government because of the overtones of autonomy (cf. §1.2.1). Some proponents argue that the shorter variant 'Kamta,' is not intended to have the political overtones of 'Kamtapuri'—the latter term associated naturally with the Kamtapur Peoples Parties. Recently the *Kāmatā Sahitya Sabhā* 'Kamta literature society' was founded. It has as one of its aims the promulgation of *Kamta* as a caste-generic, and politically inclusive, language name.

Kamrupa: Chatterjee (1926) uses this term to refer to the stage of linguistic history which is ancestral to both Asamiya and KRDS. In this study, Kamrupa is used with the same meaning, and it is therefore not considered synonymous with KRDS, which is a distinct historical stage (cf. §7.3.4).

A distinct position is put forward by Nirmal Das (2001), who maintains that 'Kamrupa' or 'Kamrupi' is a more fitting title than 'Kamta' for the KRDS varieties. This view is problematic, however, because the term 'Kamrupi' is most popularly used today to denote not KRDS, but the western dialect of Asamiya spoken in the Kamrup region of Assam (cf. U. Goswami 1970).

A further concern regarding the term 'Kamrupa' is raised by Wilde (2008:2, citing Matisoff 1999):

> The name "Kamrupa" (or, "Kamarupan"), which derives from the name of the mediaeval kingdom Kāmarūpa, is also used for the Tibeto-Burman languages of "Northeast India and adjacent areas of Western Burma, Bangladesh, and Tibet" (Matisoff 1999:173). For discussion amongst Tibeto-Burman scholars on the issue, refer to Burling (1999) and Matisoff (1999).

The potential ambiguity with Tibeto-Burman studies notwithstanding, I continue to refer to the western dialect of Asamiya as *Kamrupi*, and the historical ancestor of proto Kamta and proto Asamiya as *proto Kamrupa* (see Figure 7.20 reproduced at the end of this chapter).

---

[3] Grierson writes: 'according to Rai Gunabhiram Baruah's Buranji, this name was given to this portion of Assam by the Ahoms to denote that it had been conquered and consequently 'the people hated the name'.' (1903–28; vol.V)

Koch Rajbanshi: This term is an extended form of the more widely used term Rajbanshi, described below. The extended form specifies the descendants or race of the *Koch* kings. This specification is pertinent as there are multiple south Asian social groups that go by the Sanskrit appellation *Raja-vangshi* 'royal race'. N. Das (2001) has criticised the term Rajbanshi for this very reason, of being too broad in its reference; the same criticism applies to other generic names including deshi bhasha 'the local language'.

Rajbanshi: This term is favoured in south east Nepal by Rajbanshi caste members, and in West Bengal by Rajbanshis who favour *linguistic* autonomy but reject the *political* overtones of 'Kamtapuri'. In Nepal there is a *Rājbanshi bhāshā prachār samiti* 'Rajbanshi language publishing society,' with an office in Bhadrapur of Jhapa district. The term Rajbanshi is criticised by Barma and others for being caste-centric, and in particular excluding the Muslim population who speak the same lect but do not identify themselves by the Hindu term 'Rajbanshi'. In addition to the caste-centricism, there is the problem of breadth of reference mentioned above—that is, linguistically unrelated social groups which also go by the name 'Rajbanshi.'

Rangpuri: This term is favoured in the Rangpur area, interchangeably with 'Bahe.' Chaudhuri (1939) prefers to use Rangpuri to Rajbanshi, as it avoids the problem of being caste-centric. However, with a sizeable number of speakers now located in two different countries to Rangpur, and with no historical reason for preferring 'Rangpuri' to 'Kamta,' it is unlikely that the term Rangpuri will catch on further afield.

Surjapuri: This term is favoured in north east Bihar and adjoining portions of Dinajpur district of West Bengal by Rajbanshis and Deshi Muslims alike. The entry in the Ethnologue (Gordon 2005) under 'Surajpuri' seems to be an orthographic variation of Surjapuri. The speakers I have mixed with invariably pronounce the language name as [surʤapuri].

Tajpuria: This term is favoured in south east Nepal among Tajpurias who reject the name Rajbanshi on the basis that it is a caste designation which does not apply to them.

In this study I have chosen to use an acronym, KRDS, to refer to the lects which go by the list of names above. The acronym is not intended as a long term solution to the various debates over naming—the responsibility for such a solution rests with the speakers themselves. The acronym KRDS stands for Kamta, Rajbanshi, Deshi, Surjapuri. These terms summarise some key differences between the sociolinguistic lenses through which speakers view their mother tongue.

Finally, it needs to be noted that the meaning of these terms will inevitably change, and over time the definitions given above may become obsolete. The social situation among speakers of KRDS is in a period of flux. Different leaders in different countries and states are calling on speakers to adhere to different language ideologies, and each ideology comes with a different language name attached. The outcome is far from determined.

1.3 'Language' and 'dialect'

Distinguishing between 'a language' and 'a dialect' is notoriously problematic for the NIA lects. The problem results from the following paradox: 'language' and 'dialect' are popularly understood to be dichotomous terms—*either* something is 'a language' *or* it is 'a dialect'—and yet the very nature of a dialect *continuum* is that internal linguistic divisions are a matter of *degree* rather than dichotomy. Clearly divergent lects may be distinguished, but they are separated by intermediary cases whose status is ambiguous. Furthermore, the

variation is not one-dimensional but involves multiple geographical, social and linguistic dimensions. Therefore, lects that are clearly divergent along one dimension (or from one analytical perspective) are from another perspective merely intermediary cases whose status is ambiguous.

There are, however, (apparently) fixed *social and political points* in the speech community, and it is these factors that traditionally determine whether something is considered 'a language' or 'a dialect'. This observation is reflected in the North American language debate over the status of Ebonics (African American Vernacular English). Concerning Ebonics, the Linguistic Society of America passed a resolution which states:

> The distinction between 'languages' and 'dialects' is usually made more on social and political grounds than on purely linguistic ones. (LSA 1997)

Chambers and Trudgill (1998) use the notions of 'autonomy' and heteronomy' to analyse how language and dialect are distinguished in the West Germanic dialect continuum. Heteronomy is the opposite of autonomy, and refers to dependence rather than independence. Chambers and Trudgill provide four sociocultural phenomena from West Germanic as illustration of sociolinguistic 'heteronomy':

(i) Speakers **consider** that they are speaking Dutch;
(ii) Speakers **read and write** in Dutch;
(iii) Standardising changes are **towards** Dutch;
(iv) Speakers **look to** written Dutch as the written variety which corresponds to their spoken varieties.

This analysis clearly demonstrates that the relation between dialects and 'a language' is primarily an *ideological* relation (cf. Enfield 2003:4). The relation is between the *ideas* held by speakers regarding the varieties they speak, and the *ideas* regarding the varieties they write. These ideas are sociocultural phenomena (rather than purely linguistic), and they determine the variety which speakers select as the medium for reading and writing.

Sociocultural ideology plays a key role in differentiating 'languages' and 'dialects', and the same is true in the KRDS situation. First, the general Bengalis—those from the south of Bengal—in large part consider KRDS to be 'a dialect of Bengali'. As this sociocultural group occupies the dominant position of power within the state of West Bengal, it is their linguistic ideology which has governed official policy. Second, a good number of KRDS speakers in West Bengal ideologically understand themselves to be speaking not 'Bengali,' but 'Kamta' (or 'Rajbanshi' or 'Deshi' depending on their political persuasion, cf. §1.2.1–1.2.2). However, because as a sociocultural group they occupy a less politically powerful position than that of the Bengalis, their linguistic ideology has made only a very minor impact on government policy. In this context, it is possible to understand why the group occupying the less powerful political position feels disenfranchised by the ideology of the powerful. On the other hand it is possible to see why Bengalis generally fail to understand the sentiments expressed by KRDS speakers. Their sociocultural ideologies differ, and these ideologies shape their account of 'language' and 'dialect'.

Similarities exist between the case of Ebonics in the USA and the case of KRDS in West Bengal. Both cases involve competing language ideologies, with a disparity of power between the two sides. The socioculturally and politically dominant group grants high status to their own mother tongue, while the less powerful group is dissatisfied with

the low social and political status accorded their mother tongue. The statements of the LSA resolution (1997) are given here with some summarising and substitution. To make the analogy clearer, 'KRDS' is substituted for 'Ebonics,' 'Bangla' for 'English,' and 'West Bengal' for 'United States'. Substituted words are underlined.

1. The variety known as 'Kamta', 'Rajbanshi', and 'Deshi Bhasha' and by other names is systematic and rule-governed like all natural speech varieties. … Consequently, characterizations of KRDS as 'slang,' 'mutant,' 'lazy,' 'defective,' 'ungrammatical,' or 'broken Bangla' are incorrect and demeaning.
2. The distinction between 'languages' and 'dialects' is usually made more on social and political grounds than on purely linguistic ones. … What is important from a *linguistic and educational* point of view is not whether KRDS is called a 'language' or a 'dialect' but rather that its systematicity be recognized.
3. There are individual and group benefits to maintaining vernacular speech varieties and there are scientific and human advantages to linguistic diversity. For those living in West Bengal there are also benefits in acquiring Standard Bangla and resources should be made available to all who aspire to mastery of Standard Bangla.
4. There is evidence from Sweden, the US, and other countries that speakers of other varieties can be aided in their learning of the standard variety by pedagogical approaches which recognize the legitimacy of the other varieties of a language. From this perspective, a recognition of the vernacular of KRDS students in teaching them Standard Bangla is linguistically and pedagogically sound.

There is, however, a significant dissimilarity between the cases of KRDS and Ebonics. Over the past two decades, KRDS speakers have developed a copious written literature in their own mother tongue; Ebonics has no such written literature. If the defining characteristic of 'a language,' as distinct from 'a dialect,' is taken to be *the existence of a written literature* (and this definition is commonly assumed in South Asia), then the growing written corpus in KRDS must have some bearing on classification. The language versus dialect issue will not feature prominently in this study, though some further relevant comments are given in the concluding chapter.

The attentive reader will have noticed that the term 'lect' has already been used in this book both for entities traditionally termed 'languages' as well as for those termed 'dialects'. This technical term is synonymous with 'linguistic variety,' and encompasses both 'language' and 'dialect' without distinguishing between the two on the basis of sociopolitical status.

1.4 Review of previous linguistic studies of KRDS

Up to the present study, there has been no in-depth, modern linguistic treatment of the KRDS lects as a whole, let alone systematic reconstruction of their history. This situation is in keeping with the general pattern of NIA research as described by Blench and Spriggs (1998:10):

> The Indo-Europeanist habit of ignoring what are strangely called 'minor languages' has resulted in a virtual lacuna in research on Indo-European languages of India with only small numbers of speakers. One of the more evident tendencies in Indo-European linguistics is to give primacy to written languages, such as Sanskrit. Thus, reconstruction

> of the Indo-Aryan languages is in terms of relating the present-day forms to attested Sanskrit (cf. Turner 1962–66) rather than subjecting the body of Indo-Aryan languages to the usual procedures of historical linguistics. The consequence has been a striking inadequacy of fieldwork to describe the more than 300 unwritten Indo-European languages spoken in the India-Pakistan region in the 1990s … The conventional practice of historical linguistics in the region is thus in a rather backward state.

Previous linguistic studies pertinent to KRDS can be divided into several categories. First, KRDS has been addressed briefly in survey volumes, most significantly in the monumental, though methodologically limited, *Linguistic Survey of India* (Grierson 1903–28). The Rajbanshis and their language are also considered briefly in Voeghlin and Robinett (1977:167), and in van Driem's (2001) survey of the Himalayan languages. (The Rajbanshi are included in van Driem's survey because of their Tibeto-Burman ancestry.) A more concentrated survey of KRDS has been undertaken recently for the purpose of studying bilingualism, intelligibility, and language use in KRDS lects (Ngwazah et al. 2006). That survey follows a similar survey undertaken for the Rajbanshi lects of Nepal (Eppele and Grimes 2001).

Into the category of survey we may also place Bandyopadhyay (1991), which is the first of a multivolume dictionary project based at the University of Calcutta, under the Education Department, Government of West Bengal. The project involves the collection and publication of data for the non-standard Indo-Aryan varieties of West Bengal. Given the geopolitical scope of the project, these lects are termed by the authors as Dialectal Bengali. The lexical entries are sorted by the Indic alphabet system, and the first volume covers entries beginning with the vowels অ /ɔ/, আ /a/, ই /i/. The first volume includes maps of the delimitation of dialects as understood by the surveyors but these have not been used in this study as the criteria for delimitation are not made explicit, and most of the test locations on the maps are unlabelled. The project includes the local lects of northern West Bengal, and thus there is some overlap with the KRDS lects. The usefulness of the dictionary for comparative work is limited, however, by the ordering system which is sorted alphabetically by lexical item rather than by reference to a proto form (the method employed by Turner (1962–66), and in Appendix A of Toulmin 2006). The accessibility of the dictionary for descriptive purposes is likewise limited, because the user cannot at present sort or filter the data in any way. This dictionary project will be of significant use for future linguistic studies if it is made available in an electronic form that may be sorted or filtered according to multiple criteria.

After survey treatments, the second category of previous studies in KRDS consists of linguistic works whose scope is limited to the description of a particular KRDS lect. This category of studies does not include systematic analysis of the broader KRDS linguistic context, or reconstruction of the history of the group as a whole. Studies in this category are: for Rangpuri, Grierson (1877) and Chaudhuri (1939); for central eastern Jhapa Rajbanshi, Wilde (2002, 2008) and Poudel (2006); for eastern Jhapa Rajbanshi, Toulmin (2002a); for Goalparia lects, Datta (1971); for an unspecified Jalpaiguri variety, Sanyal (1965); and for the Cooch Behari variety, see Joshy and Joshy (2007), and Toulmin (in press). The majority of these studies are not reference grammars, the exception being Wilde (2008) and to a lesser extent Poudel (2006) and Joshy and Joshy (2007). The NIA lects immediately adjacent to KRDS in the NIA continuum are Rajshahi Bangla (described in Chaudhuri 1940, and in greater detail in Islam 1992) and Kamrupi Asamiya (described in U. Goswami 1970).

Third, KRDS has figured at the periphery of some major historical studies of other NIA lects. Most notable of course is Chatterji (1926), and then Kakati (1962), Shahidullah (1966), and Maniruzzaman (1977).

Fourth, there are studies of KRDS lects which have been undertaken within a traditional Indic or Sanskritic model of analysis (Barma 1991, 2000). The categorisation of KRDS lects as western, central and eastern (which is verified in the present study) was first put forward by Barma (1991).

Fifth, there are studies such as that by Das (1990) for Goalparia which use what I term an 'etymological method'. This method involves comparison of contemporary forms with the putative ancestral forms in Sanskrit, resulting in pseudo correspondences. The correspondences are 'pseudo' because the reconstruction is not controlled by the principle of regularity of sound change. In order to distance themselves from this pseudo comparative method, studies such as Southworth (1958), Pattanayak (1966) and Maniruzzaman (1977) refer to the conventional Comparative Method of historical linguistics as 'controlled reconstruction'.

Sixth, there are essays in N. Das (2001) and Bhakat (2004) which cover a range of sociolinguistic, linguistic and other historical topics concerning KRDS.

Seventh, there are word lists published for some KRDS varieties. Most notable is that found in Hodgson (1880), which (based on linguistic features) seems to record a western Jalpaiguri or possibly south eastern Darjeeling variety of KRDS, labelled by him as 'Koch'. U. Goswami (1974) contains a comparative wordlist of Goalparia and Kamrupi lects, and Damant (1873) gives a short list of words belonging to the Indo-Aryan dialect of the 'Palis' (pronounced Polis, and also called Poliyas) which he is unable to derive from an Aryan source. 'Poliya' denotes a Hindu social group of Dinajpur, and is used primarily on the Bangladeshi side of the border.

In conclusion to this section: despite considerable linguistic research on the standardised eastern NIA languages that border KRDS—Bangla and Asamiya—nothing close to the same degree of analysis has been undertaken for KRDS. The descriptive study of KRDS up to the present has either lacked linguistic systematicity, depth of analysis, or breadth of scope at the level of the subgroup.

In the area of historical reconstruction, the present position of Indo-Aryan studies is inadequate *even for the standardised lects*, let alone KRDS:

> Within the Eastern Indic language family the history of the separation of Bangla from Oriya, Assamese, and the languages of Bihar remains to be worked out carefully. Scholars do not yet agree on criteria for deciding if certain tenth century AD texts were in a Bangla already distinguishable from the other languages, or marked a stage at which Eastern Indic had not finished differentiating. Such agreement may emerge once the contemporary enterprise of producing serious descriptions of the modern languages has achieved its objectives. The priorities may then permit greater attention to the unfinished task of drawing rigorous maps of the past (P. Dasgupta 2003:352).

In making a fresh start on the historical study of KRDS, this study employs theories and methods which are suited to the dialect continuum context and 'drawing rigorous maps of the past' (see further Chapter 3).

1.5 Scope of the linguistic reconstruction

This study is concerned with the origin and development of the KRDS lects, but the approach taken is dissimilar to that of Chatterji's *Origin and development of the Bengali*

language (1926). Chatterji traces the linguistic ancestry of Bangla as far back in history as possible—through various stages of 'Bangla', to Middle and Old Indo-Aryan, and from there, to proto Indo-European. Consequently Chatterji's work has been described as 'not only … a complete picture of the development of Bengali from OIA, but also … a short survey of the historical development of Indo-Aryan languages from the old stage to the new' (Maniruzzaman 1977:32–33). Such an extended historical scope of reconstruction is not attempted in the present work.

This book presents a detailed account of the origin and development of KRDS since the proto Kamta period. The proto Kamta period refers to that period in history when unique KRDS linguistic features were propagated within a speech community (c. AD 1250–1550, see §7.3.1). These unique protofeatures have been inherited into the present-day KRDS lects and identify them as a subgroup distinct from neighbouring lects including Bangla and Asamiya.

It is generally accepted that KRDS forms part of a higher level subgroup with Bangla, Asamiya and Oriya (along with several other smaller lects). Chatterji (1926) termed this subgroup *eastern Magadhan* (a descendant of common Magadhan), with the Bihari lects making up western and central Magadhan. However, the eastern Magadhan subgrouping may not be as robust as previously thought (see the discussion in §7.3.2). Until further reconstruction is undertaken (1) at an all Magadhan level and (2) based on robust historical methodology—in particular, by distinguishing innovation from retention—the intervening stages between proto Magadhan and proto Kamta remain in doubt.

The present study includes phonological, morphological and sociohistorical reconstruction, which come in Chapters 4, 5–6, and 7 respectively. Reconstruction is limited to (i) the phoneme inventory for proto Kamta; (ii) inflectional morphology of proto Kamta; and (iii) the formal characteristics of some proto Kamta vocabulary. Semantic and syntactic changes are not reconstructed, except as they impinge on the construction of cognate sets or the reconstruction of morphological changes. The protovocabulary resultant from the phonological reconstruction can be found as a comparative wordlist in Appendix A of Toulmin (2006). Further limitations of this study are:

1. Exhaustive research of ancient documents in the Cooch Behar archives has not been undertaken. Such analysis may shed further light on the linguistic history of, at least, central KRDS. The present reconstruction relies primarily on spoken rather than written lects, though written lects nonetheless retain a role in establishing chronology (see §3.4.3.2). A thorough description of the use of innovative KRDS features in the historical literature remains to be undertaken. Upon completion of that task, the conclusions outlined in this study regarding the sequencing of changes may require revision.
2. This work is also not a reference grammar for any KRDS lect. A reference grammar of the socially important lect spoken in Cooch Behar remains an outstanding need. (Joshy and Joshy [2007] should be seen as a step in this direction, with Barma [in prep.] to follow soon.)
3. Morphological and morpho-phonological changes in derivational morphology (e.g. agentive nominalisation strategies) have not been reconstructed, and await further study. For further discussion of limitations to the morphological reconstruction see §5.1 and §6.1.

With linguistic study of KRDS still very much in its infancy, it has not been possible to do justice to all these areas of potential research. However, I hope that further scholarly activity will test the findings of this study against written records, and against a more exhaustive set of linguistic features including derivational morphology.

1.6 Summary of the historical findings of this study

The historical findings of this study are summarised in this section as an aid to understanding how the reconstruction of numerous details in Chapters 4–6 fits within the overall account of linguistic history in Chapter 7.

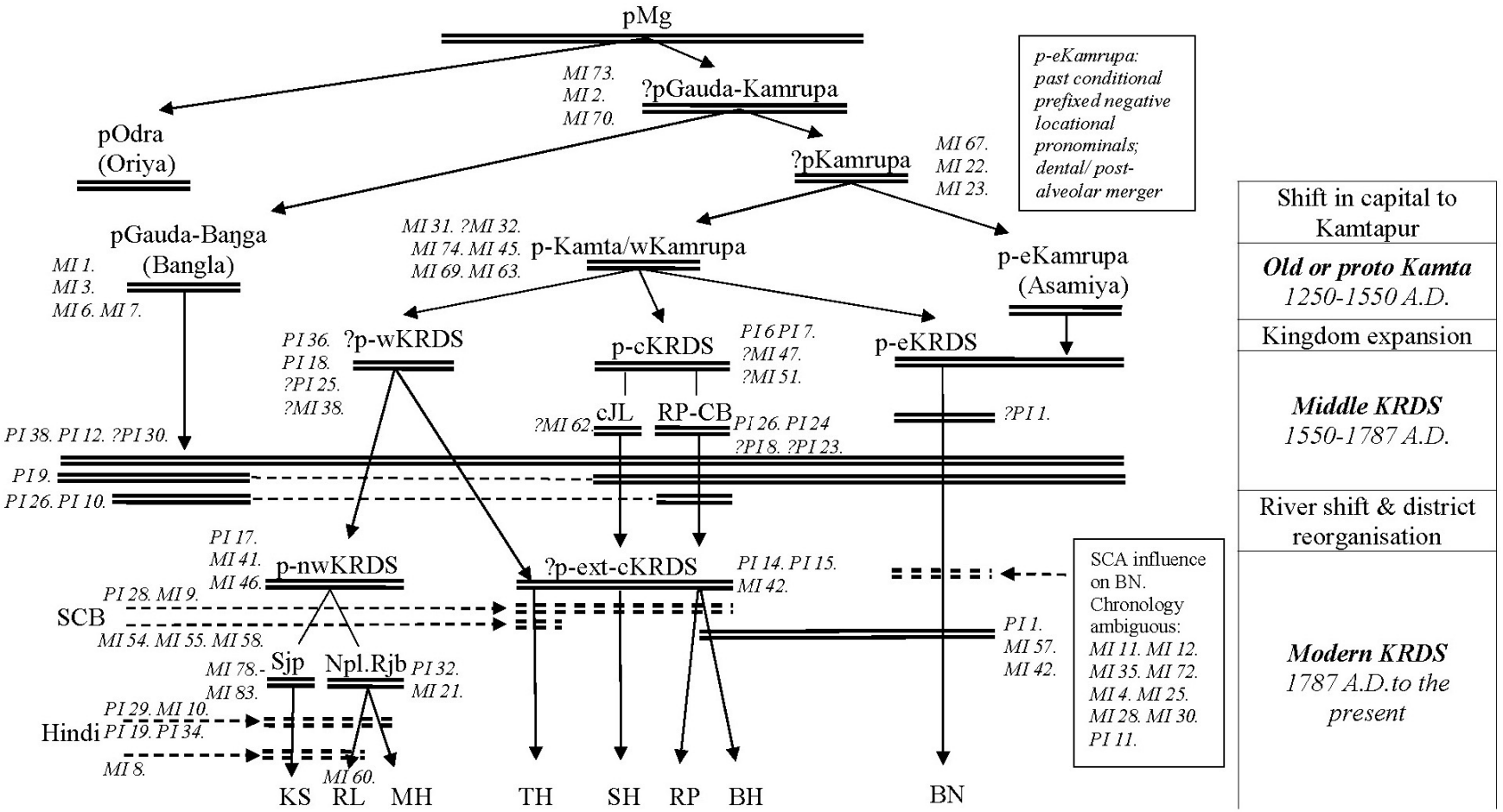
Morphological reconstruction in Chapters 5–6 provides diagnostic evidence for a common historical stage ancestral to the eight KRDS lects compared in those chapters. On sociohistorical grounds, this stage is termed 'proto Kamta' in Chapter 7 and assigned the chronology of approximately AD 1250–1550: sandwiched between the establishment of the Kamrupa capital at Kamtapur in 1250, and the political (and plausibly linguistic) expansion under Koch King Nara Narayana in 1550 (see further §7.3.1). The absence of phonological changes to define this period is not odd in its regional context. The phonologies of Bangla and Oriya were also stable during the same period, and Oriya phonology has been remarkably stable from that time to the present day.

A mix of more localised phonological and morphological innovations occurred subsequent to the breakup of the proto Kamta speech community. Each of these innovations has been assigned to either the middle KRDS or the modern KRDS period based on its geographical range and the associated methods outlined in Chapter 3. The sociohistorical events which define the historical boundary between middle and modern KRDS are the shift in course of the river Tista in AD 1787 and the reorganisation of districts under the new colonial powers at around the same time.

In addition to these localised innovations, the middle and modern KRDS stages are characterised by the propagation of innovations with a wider scope beyond KRDS. During the middle KRDS period such changes were phonological, and included the loss of final *ɔ (cf. §4.4.11) as well as changes in the voicing quality of sonorants (cf. §4.3.4). During the modern KRDS period various phonological and morphological features have entered KRDS lects due to increased diglossia with standardised State languages. In today's Bengal the influence is from SCB, in Assam it is from Asamiya, and for the Nepal Rajbanshi and Bihar Surjapuri the influence comes from Hindi, and to a lesser extent Nepali.

Special mention may be made of eastern KRDS, whose history is problematic because of its mixed linguistic ancestry. In Chapter 4, Bongaigaon (BN) is shown to have undergone the same phonological restructuring as Asamiya. In Chapters 5 and 6, BN is shown to have inherited some of the proto Kamta changes. However, where these proto Kamta features clash with proto Asamiya features, the Asamiya ancestry wins out. In many cases, it has not been possible to provide a relative sequencing of eastern KRDS's participation in proto Kamta changes as against proto Asamiya changes. This is not to say that eastern KRDS is *only* a mix of proto Asamiya and proto Kamta features—it also possesses some unique innovative features of its own (cf. §4.3.1 and §5.4.1).

This summary of the reconstructed history is modelled in Figure 7.20 (reproduced below) using a tree diagram which has been schematically altered in keeping with the sociohistorical theory of language change outlined in Chapter 3.



Reproduction of **Figure 7.20:** The linguistic history of KRDS from proto Magadhan, through proto Kamta, middle and modern KRDS, to the present

2 *Research design*

The central design components for this study are its two stage methodology, and the selection of variables and interpretive methods for each stage.[1] 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.

2.1.1 Stage 1: more linguistic data, less informants

The first stage began with the collection of large amounts of linguistic data at eight collection sites across the KRDS area. More than 1300 lexical and morphological items were collected at each site, and these are available as a comparative wordlist (Appendix A of Toulmin 2006). The wordlists are transcribed phonemically, drawing on the phonological analyses of previous studies (see §1.4) and fresh analysis where necessary. The phonologies of KRDS lects can be gathered from the descriptive sections of Chapter 4. The data which underpin the comparative reconstruction thus adhere to Southworth's criterion (1958:56) 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.'

Linguistically, this reconstruction focuses on the phonological form of reconstructed lexical items and morphemes. Detailed reconstruction of meaning changes typically occurs subsequent to the reconstruction of forms, and has not been undertaken in this study. Consequently, the data (Appendix A of Toulmin 2006) should be viewed as a precise representation of etymological *forms*, with the English glosses provided as approximations of the meaning of each item. Despite being approximations, the meanings given are judged to be accurate enough to justify cognacy (which is to the point for this study).

[1] A full discussion of the theory and methods which guide the interpretation of data is found in Chapter 3.

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

- Non-animate nature;
- Animals;
- Human body;
- Human food;
- Human experiences/emotions;
- Human products;
- Human relationships and kinship terms;
- Human jobs;
- Village household terms;
- Parts of the banana tree (due to its cultural importance);
- Religious terms;
- Diseases, disabilities, illnesses;
- Postpositions, conjunctions;
- Numbers;
- Days of the week, months of the year;
- Adjectives;
- Verbs, including causatives.

The morphological items which were collected include the following sets:

- Nominal inflections of case, noun class and number;
- Personal pronouns;
- Pronominals of quantity (e.g. 'this many') and quality (e.g. 'like this');
- Temporal and locational pronominals (e.g. '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 eight collection sites, whose parents had spoken the local lect as their mother tongue, and who had sufficient linguistic interest in their mother tongue to sustain data elicitation over several days. Biodata for the informants are given in Appendix C of Toulmin (2006).

The location of each of the test sites was selected, firstly, on a geographical basis, with an approximately equal distance between each site. Secondarily, 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, and 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 the Government of Nepal.
- 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 is highly similar to the lect spoken over the border in the south west corner of Darjeeling district, India (Toulmin 2002b).

- 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. §1.2.2).
- Shalkumar, located in the centre of Jalpaiguri district, India. Speakers in this area refer to their mother tongue variably as 'Kamta(puri)', 'Rajbanshi' or 'Deshi bhasha' as described in §1.2.2
- Rangpur, the headquarters of a district in Bangladesh. During this first stage of research, data were collected with speakers at several sites outside the town perimeter (cf. Appendix C of Toulmin 2006). 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 with Cooch Behar district. The second town is located 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. Data were collected with speakers residing in old Bongaigaon town.

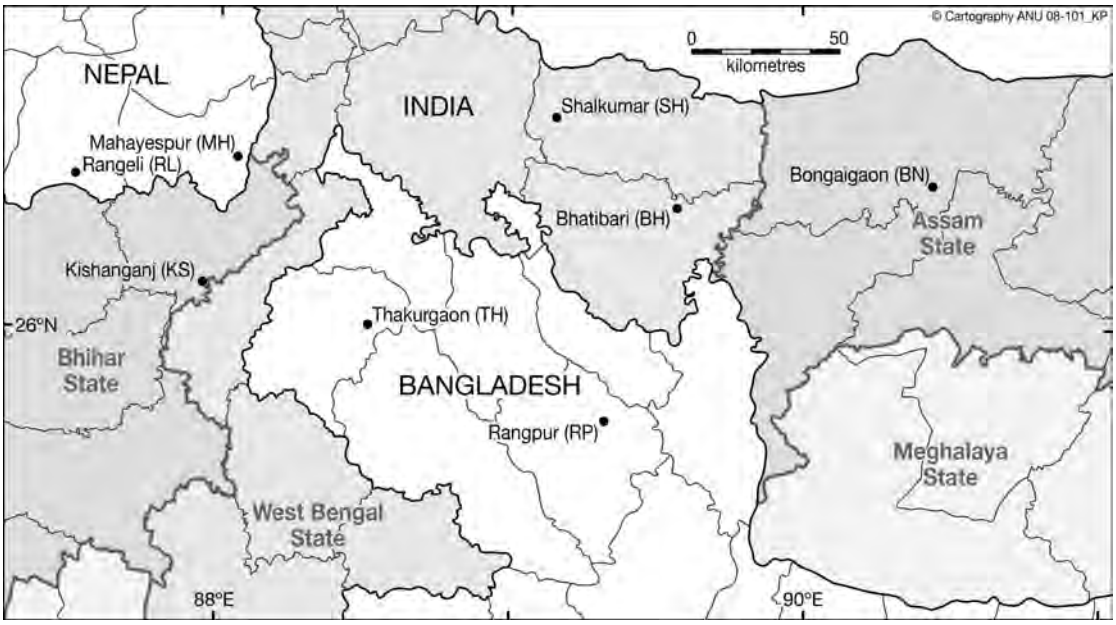


Figure 2.1: Location of sites for the first stage of data collection

The data collected during Stage 1 were analysed using the Comparative Method of historical linguistics. Methodology for phonological reconstruction is discussed in Chapter 3. Regarding 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 sociolinguistic reconstruction that results in Chapter 7.

2.1.2 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, and 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 innovations.
- Diagnostic value of the change: the linguistic variables tested were those reconstructed as diagnostic of propagation events (cf. §3.4.1), or otherwise useful for the sociolinguistic reconstruction in Chapter 7.

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

- Geography: the collection sites were located much closer together than for Stage 1, with a distance of 25–30 kilometres separating each site. The geographical locations are shown in Figure 2.2 with each location given a unique numerical code. The names of the village or town at which data were collected are given in Appendix D of Toulmin (2006).

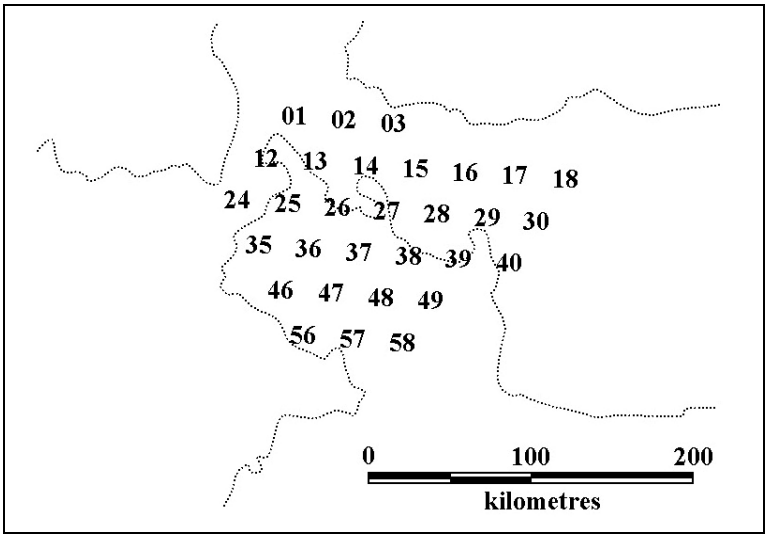


Figure 2.2: Geographical location of the test sites for Stage 2 data collection

- In the case of Site 24, and to a lesser extent at some other sites, Muslims and Hindus were found to be living in close proximity. Several speakers were interviewed from each of the two socioreligious communities, though not all the data have been analysed as yet.

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 five 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 using one of the names listed in §1.2.2.
- Education level: informants for the collection of lexical and nominal morphological data were either uneducated or minimally educated (up to around five years). For collection of verbal morphological paradigms, more educated speakers were required who (usually) were more capable of keeping one grammatical category constant (tense-aspect) while changing another category (grammatical subject): 'I do, you do, he does, etc. etc.'
- Mobility/stability: each informant had not lived away from the village for more than five years.
- Age: wherever possible, each informant was between 30 and 55 years old.

Data collection for this second stage was carried out entirely monolingually in KRDS, with pictures assisting the 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 with respect to gender is not part of this study, because in general it does not affect the historical reconstruction. 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. The results of the pilot concur with Toulmin (2002b). In that study 110 KRDS speakers at 11 villages across northern West Bengal and Jhapa district of Nepal were interviewed regarding their perception of linguistic variation in KRDS. In none of those interviews did participants suggest that sex was a significant factor for linguistic variation in their speech community. 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 interviews conducted for the present study, little variation was observed along sex based lines. The one exception was recorded at Site 17 (near Gosaigaon in Assam), where progressive raising of *a was present for female speakers but absent for male speakers (see Appendix D of Toulmin 2006).
- Attitude towards mother tongue: With language/dialect status a politicised issue in West Bengal and Assam, questions about a speaker's language attitudes were considered too threatening and stressful for the informants—putting the reliability of their answers 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 a relatively 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 four or more speakers interviewed at almost all sites. The schedule of biodata questions and the concepts elicited in these interviews are given in Appendix E of Toulmin (2006). The results of this dialectological research are given in full in Appendix D of Toulmin (2006), and included in the historical argument laid out in Chapter 7 of this book.

3 *Theory and method of reconstruction*

3.1 Introduction

The goal of this chapter is to articulate a reliable method for reconstructing linguistic history in a dialect continuum. Desiderata for such a method are outlined in §3.1.1, followed in §3.1.2 by an introduction to the characteristics of a dialect continuum which make historical reconstruction problematic.

The method proposed in this chapter is based on a sociohistorical theory of language change (see §3.2). The sociohistorical theory guides the criticism of the traditional Comparative Method (§3.3) and the development of a revised method (§3.4). The more innovative aspects of the revised method concern the *sociohistorical conditioning of language change*. Reconstruction of this sociohistorical conditioning is guided by a sociohistorical linguistic typology (§3.5). This total methodological and theoretical framework is put to the test of reconstructing KRDS linguistic history in Chapters 4–7.

3.1.1 Desiderata for a method of historical reconstruction

The desiderata outlined here are a summary of some of the principles which modern linguists bring to the task of historical reconstruction. The framework presented in the rest of this chapter is then an argument regarding how these desiderata might be satisfied for the history of a dialect continuum.

Our methods of historical reconstruction should, in the maximum number of contexts, have the capacity to:

1. Reconstruct change events in linguistic history, that is, innovations.
2. Reconstruct the historical sequencing of these change events.
3. Reconstruct continuous transmission in linguistic history, that is, linguistic inheritance.
4. Reconstruct the time depth for continuous transmission of a linguistic feature. This time depth may be in terms of either ‘absolute’ time, or more commonly, ‘relative’ to the time depth of other features.

5. Reconstruct the historical relations between related lects, that is, linguistic phylogeny. Such relations are based on continuously transmitted material inherited from a single prior change event.
6. As a consequence, reconstruct the linguistic material (words, morphemes, phonemes) present at every stage in linguistic history.

These desiderata set out an ideal which reconstruction methodology strives towards. However, the central method of historical linguistics—the Comparative Method—when used alone, is not equal to the task of fulfilling these desiderata for dialect continua (as shall be argued in §3.3). The next section explores the sociolinguistic characteristics of a dialect continuum which make historical reconstruction problematic.

3.1.2 The dialect continuum

New Indo-Aryan is commonly characterised as a dialect continuum. Masica describes the situation:

> Lacking clearcut geographical units of the European type where dialectal variants can crystallise in semi-isolation, or longstanding political boundaries, the entire Indo-Aryan realm (except for Sinhalese) constitutes one enormous dialect continuum, where continued contact inhibits such crystallisation, and differentiated dialects continue to influence one another. The speech of each village differs slightly from the next, without loss of mutual intelligibility, all the way from Assam to Afghanistan. (Masica 1991:25; cf. also Shapiro and Schiffman 1981)

The dialect continuum is a sociohistorical and sociolinguistic configuration of speakers and their lects, characterised by non-discrete boundaries between communities of speakers. Lects in a dialect continuum are closely related; their histories are characterised by recurrent interaction; and the barriers to communication between speakers of adjacent lects are weak or non-persistent. There is a correlation between linguistic divergence and geographical distance (or, in the case of social dialect continua, social 'distance').

Concurrent with *localising* tendencies towards greater divergence, there are *integrating* social forces which encourage linguistic unification through dialect levelling. Saussure (1966:281) termed these two forces '*l'esprit de clocher*' ('spirit of the (town) church spire') and '*la force d'intercourse*' ('force of communication').[1] To use a different metaphor, both centrifugal ('centre fleeing') and centripetal ('centre seeking') forces act concurrently in a dialect continuum. The forces involved are social in nature and pull the speaker's linguistic conventions in two directions at one and the same time. The centripetal forces pull the speaker's conventions in the direction of more inclusive and 'global' social norms. These forces support the propagation of innovations across a wider range, and support the maintenance of features shared within a wider community of speakers. The centrifugal forces move in the opposite direction, towards more localised and exclusive usage. These forces support the maintenance of conservative local features (even in the face of wider pressures to replace them) as well as the propagation of innovative local features.

---

[1] 'Dans toute masse humaine deux forces agissent sans cesse simultanément et en sens contraires: d'une part l'esprit particulariste, l''esprit de clocher'; de l'autre, la force d''intercourse', qui crée les communications entre les hommes' (Saussure 1966:281).

Given this depiction of the dialect continuum, the possibility of common changes occurring subsequent to a period of divergence emerges as a completely normal possibility. This possibility applies equally to the histories of standardised 'languages' within the continuum, as to the histories of minor lects. If our reconstruction method is to be equal to the challenge of dialect continua, then it must reconstruct *all* possible change events, and not treat normal change phenomena as irregular or irrelevant.

3.2 A non-essentialist, sociohistorical theory of language change

3.2.1 Overview

The framework of historical reconstruction outlined in this chapter is based on a sociohistorical theory of language change (following Weinrich, Labov and Herzog 1968; J. Milroy 1992; Croft 2000; Enfield 2003).[2] Within this framework, **linguistic phylogeny** is defined by the *differential transmission of linguistic material through speaker interaction over the course of history*. Speaker interaction enables **propagation** of novel variants, by means of social networks of speaker interaction (Croft 2000:8, following the work of J. Milroy and L. Milroy). When an innovative feature is propagated through a web of interconnected networks of speaker interaction, this is termed a **Propagation Event** (PE). The geographical and social extent of a PE is termed its **range**.

The theoretical accompaniments of a PE are the **Propagation Network** (the population of interacting speakers who participated in the change), and a **Propagation Defined Language** (the population of utterances used by the members of the Propagation Network, cf. Croft 2000:26). The propagation defined language (henceforth PDL) *does not necessarily represent a structurally homogenous entity, because it is not defined by its overall linguistic homogeneity but by a PE*. A phylogenetic **subgroup** of lects share a common PDL (defined by a PE) as part of their linguistic phylogenies. The network of propagation may have been smaller than for earlier PNs (denoting a phylogenetic division), or larger than for earlier PNs (denoting a phylogenetic reticulation). Whether by division or reticulation, the new PN defines a new PDL which will figure in the phylogenies of any descendent lects.

It seems to me neither necessary, nor helpful, to distinguish terminologically between PDLs that result from phylogenetic reticulation and those that result from phylogenetic division. Accordingly I take it that *phylogenetic reticulations define a subgroup of lects in the same way as do phylogenetic divisions.* The theoretical Maxim 'once a subgroup, always a subgroup' does not generally hold in a dialect continuum, and a non-essentialist theory of linguistic phylogeny does not insist that it should hold.

A crucial task of the linguistic historian is the **sequencing** of PEs. Sequencing enables innovations to be interpreted within a coherent and plausible account of **linguistic history**. Three criteria for sequencing linguistic changes are put forward in this chapter: **linguistic**, **textual**, and **sociohistorical**. The first criterion concerns linguistically necessary (or plausible) seriation of changes—such as when one change 'bleeds' the linguistic conditions for another. The second criterion for sequencing involves using evidence from historical texts. The third criterion is the most innovative of this study, and involves using disjunctions between PNs as evidence for an intermediary change in the structure of the

[2] In the following summary statement, key terms are marked in bold, and detailed discussion of their definitions is left to subsequent sections.

speech community—a **Speech Community Event** (SCE). The reconstruction of SCEs then informs the sequencing of PEs. Sociohistorical linguistic seriation is guided by **sociohistorical linguistic typology**, just as linguistic seriation is guided by **diachronic linguistic typology**.

3.2.2 A sociohistorical theory of linguistic phylogenesis

The idea of phylogenesis is central to the discipline of historical linguistics and its task of reconstructing linguistic history. The general idea comes of course from biology, where it provides an explanation for continuities and discontinuities between species in terms of historical origins and subsequent evolutionary developments. The notion of 'a species' is at the heart and centre of the biological theory of phylogenesis—it shapes how evolutionary history is construed and reconstructed. The parallel notion for a linguistic theory of phylogenesis is the definition of 'a language', or 'a lect'. The definition of this notion has a similar effect on how we construe and reconstruct linguistic history.

For the purposes of historical reconstruction, the definition of 'a language' must be informed by an understanding of the mechanisms by which 'a language' changes. One of the tasks in this section is to articulate a definition of 'a language' which (a) conforms with current theories of language change, and (b) is useful for the purposes of historical reconstruction.

The notion of 'a language' may be defined differently given different purposes. The descriptivist may define a language for her purposes using synchronic, structural criteria. For the purposes of historical reconstruction a different definition may be required. Croft (2000) compares two kinds of definition that may be taken up by language change theory—the essentialist definition, and the population-based definition. The essentialist definition of a language is constructed in terms of essential structural features. Chatterji attempts such a definition for the 'Bengali' dialects by listing a set of features 'that may be called pan-Bengali' (1926:139).

Whatever the descriptive value of this kind of definition, it is theoretically highly problematic when applied to historical reconstruction—a point which Croft explains using biological evolutionary theory:

> In the ESSENTIALIST view of a species, each species has immutable essential structural properties that identify it (Mayr 1982). … The essentialist view ran into problems due to various sorts of structural variation among species, including high degrees of structural variation among individuals in a population and also among different life-stages in an individual in a population … The essentialist view also ran into problems with populations which could not be distinguished by structural features but were distinct reproductive communities … But the greatest problem for the essentialist view of a species is that a species evolves, and in so evolving, can lose 'essential' structural properties. Identifying this problem is one of the major contributions of Darwin to evolutionary biology. (Croft 2000:13)

Linguistic analogies to these problems are: (1) the existence of considerable structural differentiation between utterances of the same language—'orderly heterogeneity' (Weinrich et al. 1968); (2) changes in the structural features of the lect used by a single speaker over the course of her lifetime; (3) the existence of speech communities that share many features but have ceased to be connected by networks of speaker interaction—for example, the Hindi spoken in India and that spoken in Fiji; and (4) the fact that a language changes and may lose those defining characteristics. (These problems all apply to Chatterji's 'pan-Bengali characteristics' mentioned above; for a detailed critique see §7.3.6).

An alternative approach to the essentialist definition of 'a language' is the population-based definition. Having abandoned an essentialist definition of 'a species', it is the population-based definition which is now adopted by evolutionary theory:

> The POPULATION theory of species is completely different from the essentialist theory (Mayr 1982:272). A species consists of a population of interbreeding individuals who are reproductively isolated from other populations. This property—interbreeding, and lack thereof between species—is the 'essential' property the individuals have in common. … This is a radically different view of the species as a conceptual category. *The category definition is based on a specific set of individuals, and category membership is defined in terms of how the individuals interact with each other, not by any specific traits associated with all and only the individuals in the category*. (Croft 2000:13, emphasis added)

The immediate linguistic implications of a population-based approach to definition are as follows. 'A language' is defined (for phylogenetic purposes) as the set of utterances produced by a population of interacting speakers. The membership of this population of speakers is defined 'in terms of how the individuals interact with each other' (ibid.). Thus, it is speaker interaction which is the defining characteristic of 'a language', not particular linguistic features per se. The linguistic characteristics of 'a language' flow out of this definition of a language, but do not define it.

Sociolinguistic heterogeneity within the speech community is not a problem under this definition of 'a language' because the heterogeneity results from the patterns of interaction between speakers. On the other hand, communities of speakers no longer interacting with each other, but still sharing similarities of linguistic structure, are no longer a single (phylogenetic) language. Phylogenetic relations between lects are altered whenever there is a change in the social patterns of interaction between speakers.

While it is speakers' *interactions* that define 'a language', the linguistic historian has no direct access to *past interactions*. Interaction patterns must be inferred from the differential presence of linguistic traits. Accordingly, linguistic phylogeny may be defined as *the study of the linguistic features which are diagnostic of historical interactions between speakers*.

3.2.3 The mechanisms of language change

Several recent studies explore the theoretical implications of adopting a non-essentialist definition of 'a language' in historical reconstruction.[3] The discussion here focuses on the models of change articulated by Croft (2000) and Enfield (2003). Though using different terminologies, they endorse models with basically the same components. Croft's model uses the terms of evolutionary theory (following Hull 1988), while Enfield's model takes its terms from epidemiology (following Sperber 1996). Both are influenced by Keller (1994) and committed to the principle of methodological individualism: 'the explanation [of a language change] is based on acting individuals, not languages, structures, processes, or collectives' (Keller 1994:121). James Milroy makes essentially the same point: 'It is *speakers*, and not *languages*, that innovate' (J. Milroy 1992:169, emphasis original).

For the purposes of reconstructing history, not only the *entities* of historical change, but also the *processes* of change must be spatiotemporally describable. Enfield requires a 'no telepathy assumption' (2003:2; citing Hutchins and Hazlehurst 1995), which entails that

---

[3] Compare Thomason and Kaufman (1988), J. Milroy (1992), Keller (1994), Croft (2000), Aikhenvald and Dixon (2001), Enfield (2003).

the hearer has *no direct access to the speaker's mind* but must make linguistic inferences based on the behaviour of the interlocuter—in particular her speech, as it is transmitted through air. This principle may seem to the reader to go without saying; however until Weinrich, Labov and Herzog (1968), mainstream historical linguists often ignored spatiotemporal mechanisms. Some contemporary theorists (e.g. Lass 1997) continue to argue that historical linguistics should ignore these mechanisms.

In the model presented by Croft (2000), language change is broken down into two basic mechanisms. In evolutionary terms they are: *altered replication* and *differential selection*. The first of these processes involves 'the creation of a novel variant by altered replication of a lingueme in an utterance' (ibid.:238). Croft terms this mechanism 'innovation' following James Milroy (1992). However, this term will not be used here, as 'this would entail a redefinition of 'innovation', a basic term in historical linguistics' (Ross 1997:256, fn.4). Croft's 'altered replication' will instead be referred to as *novel variation*. This process is a necessary but not sufficient condition for language change. That is, novel variation does not ensure language change, but language change requires that novel variation must have occurred. A similar point is made by Weinreich et al.: 'Not all variability and heterogeneity in language structure involves change, but all change involves variability and heterogeneity' (1968:188). The causal mechanisms for novel variation are functional and linguistic, rather than social (Croft 2000:8).

'Differential selection' entails the increase in frequency of a novel variant in speakers' utterances. Croft and Enfield refer to this mechanism as 'propagation', a term which is adopted in the methodology of this study.[4] The causal mechanisms of propagation are social, not functional (Croft 2000:8). J. Milroy (1992:169) refers to this second process simply as 'change'. Just as 'novel variation' is preferred over 'innovation' to denote the first process, the more restricted term 'propagation' is preferred over the generic term 'change'.

Unlike novel variation, propagation is a sufficient condition for language change—the propagation of a novel variant in the speech community guarantees a change in the language. However, Croft leaves open the possibility that language change may occur without propagation; that is, it is not a necessary cause for *all types of structurally similar change*. 'Drift' is a type of evolutionary change defined as 'a shift in gene (lingueme) frequencies that occurs through altered replication but without selection [i.e. novel variation without propagation—MT]' (Croft 2000:235; following Hull 1988). Linguistic phylogeny has been defined above as the study of the linguistic features which are diagnostic of *speakers' past and present interactions*. Drift occurs apart from propagation and is therefore not diagnostic of speakers' interactions. Consequently, drift is *of no phylogenetic significance*. Propagation may not be a necessary condition for *all* types of language change, but it is a necessary condition for *all phylogenetically significant language changes*.

Despite the possibility of drift, sociolinguists have argued that the use of linguistic variants by a population of speakers is generally not random, but is *socially structured*. This generalisation has been empirically verified by numerous sociolinguistic studies following the seminal work of Labov (1965, 1966). Weinrich et al. (1968:187) summarise the theoretical implications of these studies in such a way that dovetails with the conclusions above:

---

[4] This process is also called 'implementation of change' (Trask 2000:159).

> Linguistic change is not to be identified with random drift proceeding from inherent variation in speech. Linguistic change begins when the generalization of a particular alternation in a given subgroup of the speech community assumes direction and takes on the character of orderly differentiation.

In the terms of this study: *only when the range of an innovative feature results from propagation is the change phylogenetically significant.*

Before moving on from the mechanisms of change, it remains to discuss the possibility of independent and parallel innovations. The first mechanism of change—novel variation—is conditioned by the structural starting conditions already present in a lect. Especially in a dialect continuum, where nearby lects share a high degree of structural similarity, it is conceivable that the same novel variant may be independently replicated on more than one occasion in distinct sections of the continuum. If these separately replicated novel variants undergo propagation, then two phylogenetic outcomes are possible. The first is that the two sociolinguistically distinct, but structurally similar, replications of novel variation undergo propagation through interconnected social networks such that they merge into a single propagation event. This scenario is diagnostic of a single propagation event, and has the same phylogenetic significance as if the novel variant had been replicated only once to start with and all other replications were the result of propagation. The second possibility is that the distinctly replicated variants undergo propagation within social networks that are not connected by speaker interaction. In this case, the final range of the innovation is not diagnostic of a single propagation event, but rather two (or more) propagation events. In this scenario, independent and parallel propagation events of a structurally similar innovation may mask phylogenetic relations—defined by single, networked, propagation events. It is the task of §3.4.1 to set out diagnostics which can filter out such changes in a principled manner.

To summarise the model: the basic components of language change are novel variation and propagation. Novel variation is a necessary condition for propagation. Propagation is a necessary and sufficient condition for *phylogenetically significant* change. If the conventional replication of linguistic features is broken through the propagation of a variant V, then a phylogenetic change event has taken place. The population of speakers using utterances characterised by V is phylogenetically distinct from all other populations of speakers, and termed a 'Propagation Network'.

The term 'Propagation Network' is intentionally chosen here, instead of 'Speech Community', for two reasons. First, 'Propagation Network' is explicit in defining the population of speakers in terms of a Propagation Event. The term SC, while widely used in historical linguistics (with more than one definition) is not explicit about *propagation*, and hence would need to be redefined with a very specific (and unconventional) sense. Second, a SC may include multiple coexisting PNs, all defined by PEs.[5] The concept of a SC is multilayered and complex; the concept of a PN is flat, defined by a PE. Both concepts are useful for historical reconstruction. The utterances produced by the PN of speakers are the 'Propagation Defined Language'. Lects that inherit linguistic material from this PDL constitute a phylogenetic subgroup (cf. §3.3.3).

We now consider the implications of this theoretical position for methodology of historical reconstruction.

---

[5] I am grateful to Dr Harold Koch who suggested this formulation of the relations between a SC and a PN.

3.3 Critical review of the Comparative Method

The Comparative Method (CM) is the most successful tool in the historical linguist's kit.[6] Yet ever since the first great successes of the method in the 19th century, the criticism has been voiced that linguistic history is often not as simple as comparative reconstruction might have us believe. This point has been argued alike by practitioners of dialect geography and sociolinguists. In a dialect continuum situation the prognosis is even worse: not only might the CM simplify linguistic history, it might distort it all together (see §3.3.1).

The results of the CM are traditionally interpreted using the essentialist notion of a homogenous protolanguage, which splits into separate daughter languages. This assumption is, in general, inappropriate for dialect continua. Can the CM be separated from essentialist definitions of 'a language' and linguistic phylogeny? This question is addressed in §3.3.2 with the implications for subgrouping theory considered in §3.3.3.

3.3.1 A case study from Indo-Aryan

Within the NIA literature, there is an excellent study by Southworth (1958) which demonstrates that (1) in this case the protophonology reconstructed by the CM is historically non-contemporaneous; and (2) the sequencing of reconstructed changes cannot be established by the CM alone. The central facts and key findings of this study are summarised below.

In his doctoral dissertation, Southworth (1958) applied the CM to a reconstruction of the historical relations between four major Indo-Aryan lects: Panjabi (Punjabi), Hindi, Bengali and Marathi. The result is a reconstructed 'protophonology' for a 'protolanguage', which he uses to reconstruct 'protowords'. However, on comparing these results with ancient written records for Middle and Old Indo-Aryan, it becomes apparent that 'the protolanguage cannot be assumed to represent any Indic dialect which could actually have existed; it combines different chronological stages … and different dialectal representations' (1958:160). The method leads Southworth to reconstruct lexical items such as *k̲ét 'field' and *g̶i̶n- 'count', which correspond to Sanskrit *kṣétra* and *gṛ̥nti* respectively.[7] The problem is that the reconstructed items combine non-contemporaneous phonological features, and thus the reconstruction is not historically accurate. In this instance, the CM failed to reconstruct reliable linguistic history. Readers who wish to skip the finer details of this case study may move on to §3.1.2.

The set of protophonemes which are reconstructed by Southworth based on recurrent correspondences includes two protophonemes labelled as *k̲ and *i̶. Southworth compares the cognates containing reflexes of these protophonemes with putatively ancestral Old Indo-Aryan forms to find that *k̲ is the mostly regular reflex of Sanskrit *kṣ* (1958:§133), and *i̶ is the mostly regular reflex of Sanskrit *ṛ̥* (1958:§148). Here 'mostly regular' means that *most but not all* of the contemporary forms which are plausibly derived from Sanskrit

---

6 For readers familiar with the NIA literature: when I speak of 'the Comparative Method' I am referring to the method of 'controlled reconstruction' as practiced by Southworth (1958), Pattanayak (1966), Maniruzzaman (1977), etc. I am not referring to the 'pseudo'-Comparative Method, discussed in §.4, which is not controlled by the principle of the regularity of sound change.

7 The symbol *ṛ̥* is used in Indic studies for 'syllabic r'. The open circle below the *r* is not to be confused with the IPA convention for 'voiceless'. Here and throughout this book, italicised Indic words are romanised written form in an actual text; phonemic forms are non-italicised and enclosed in slash markers.

forms containing *kṣ* and *r̥*, are reconstructed with protophonemes *k̲ and *ɨ in the positions that correspond to Sanskrit *kṣ* and *r̥*, respectively.

A small number of etyma with phonemes that correspond to Sanskrit *kṣ* are reconstructed as *c̲ instead of the expected *k̲, and a small number of Sanskrit forms with *r̥* are reconstructed with *i, *a, or *u, rather than *ɨ. Southworth argues that this partial mismatch between the written Sanskrit forms and the reconstructed protophonemes is the result of mixing during the protostage between (A) dialects maintaining the phonemic distinctions of Sanskrit *kṣ* and *r̥* and (B) dialects that had lost these distinctions by merging Sanskrit *kṣ* with *čh*, and *r̥* with certain other vowels. This hypothesis of borrowing to explain the irregularity in correspondences is a fair one. The main point is that, *on the whole*, the reconstructed protolanguage represents a lect which had *retained* the phonemic distinctions corresponding to Sanskrit *r̥* (*ɨ) and *kṣ* (*k̲). These protophonemes are accordingly included in the reconstructed protowords, for example: *k̲ét 'field' and *gɨn- 'count'.

The presence of these phonemic distinctions in the protolanguage causes major problems in dating the protolanguage. Sanskrit *kṣ* was lost as a distinct phoneme through merger with other phonemes prior to the 3rd century BC—except in the Kashmir area where it is retained as a distinct phoneme up to the present day. Evidence for the dating of this merger comes from the Asokan edicts—inscriptions of the 3rd century BC—which record the regional Middle Indo-Aryan dialects of the time. Southworth concludes that in order for the CM-reconstructed parent language of Panjabi, Hindi, Bengali and Marathi to include a distinct protophoneme *k̲ (Sanskrit *kṣ*) the protolanguage must represent *a language spoken prior to the 3rd century BC* (1958:§163). Furthermore, Sanskrit *r̥* (which corresponds in most cases to protophoneme *ɨ) was lost as a distinct phoneme 'in all later dialects' (presumably later than Sanskrit, including all MIA lects), with a few sporadic cases of merger even in the earliest Sanskrit. In order to account for the reconstruction of *ɨ (Sanskrit *r̥*) as a distinct protophoneme, Southworth says that the parent speech of Panjabi, Hindi, Bengali and Marathi must be dated *considerably earlier even than the 3rd century BC chronology suggested for *k̲ above.*

In contrast with the protophonemes whose presence indicates an early MIA or OIA protolanguage, Southworth shows that the protolanguage also includes features which *are much more recent innovations*. The protolanguage represents a lect which (1) lacks the Sanskrit final vowels (1958:§131), (2) lacks the Sanskrit single intervocalic stops (1958:§134), and contains assimilated and reduced counterparts of the Sanskrit intervocalic stop clusters (1958:§132). The first of these characteristics—the absence (or in some cases coalescence) of final vowels—was the result of a decidedly 'New' (or 'modern') Indo-Aryan change (1958:§160). The OIA final vowels were retained at least up to the time of the Prakrit grammarian Hemacandra, who was born in Gujarat in AD 1089. In many NIA lects these vowels were retained even up to the 16th century (cf. Masica 1991:196, and §4.4.11 of this study).

The second set of protolanguage features which complicate dating are the reflexes of OIA medial stops. The single intervocalic stops had been lost by the time of the protolanguage, and the clusters had generally been reduced (Southworth 1958:137–138). However, in the MIA literature the OIA stops are unchanged as late as Aśvaghoṣa's dramas of the 2nd century (Southworth 1958:155).

In summary, the 'protolanguage' reconstructed by the CM incorporates: features lost before MIA (*r̥* =*~~i~~), features lost during early MIA (*kṣ*=*k̲), features innovated later in MIA (changes to medial stops), as well as features innovated during NIA (loss or coalescence of final vowels). The result is that protowords reconstructed by the CM (for example: *k̲ét 'field' and *g~~i~~n- 'count') are historically inaccurate and unreliable: they combine non-contemporaneous phonological features.

The implications of this failure of the CM are considerable. *Southworth only knew the reconstructed protolanguage was historically fallacious because he had written records with which to compare his reconstruction*. In the case of KRDS, and indeed of most NIA lects, no or few written records are available to us. Is the CM still useable? Can its results be interpreted in a way that avoids the kind of historical distortions illustrated above?

3.3.2 Cutting away essentialism from the Comparative Method

It is not the intention of this chapter to throw the baby of the CM out with the bathwater of essentialism. One failure of a method does not invalidate its many successes. Nonetheless, changes must be made in how we apply the classical CM if we are to satisfy the desiderata outlined in §3.1.1.

Traditional assumptions of uniform protolanguages and discrete phylogenetic divisions between languages have failed in some, but not all cases. For example, the CM has been applied with great success in reconstructing the history of the Austronesian language family. In many cases the protophonemes reconstructed from Austronesian phonological correspondences can be perspicuously interpreted in terms of interstage languages interrupted by discrete splitting events (cf. Pawley and Ross 1993). However, even within Austronesian there are linguistic histories which require other concepts and kinds of genetic relations in order to make sense of the correspondences (cf. e.g. Ross 1997, 1998). Most notable in this respect is Geraghty's (1983:277) conclusion about how to interpret the Fijian correspondences in terms of historical events.

> Exclusively shared features merely serve to suggest that languages were once in contact, and if features are shared exclusively by languages which are not in contact, those features constitute strong evidence that the languages were once in contact. In a dialect chain such as exists in Fiji, however, all adjoining communalects have generally maintained some degree of contact, so any observed innovation can be attributed to any time between the establishment of the dialect chain and the present. *A feature found all over Fiji, therefore, may be a recent innovation.* (my emphasis)

This historical finding complicates historical reconstruction, because the correspondences generated by phylogenetic splitting may look identical to those generated by phylogenetic reticulation after an earlier split. Some linguists have responded to this challenge with pessimism:

> It is a well-known axiom that the Comparative Method is powerless if two (or more) languages undergo the same change after split-off point. … What makes the Comparative Method work is that different languages usually undergo different changes. (Anttila 1989:252)

This statement is an important caution, but overstates the point. The CM can be defined in a broad as well as a more narrow sense, and they are not both equally impotent in the

face of common innovations after differentiation. A broad definition of the CM includes the following components:[8]

CM (broad definition)

(1) construct correspondences between phonemes in putative cognate sets;

(2) reconstruct protophonemes and subsequent changes which account for the correspondences as regular phonological reflexes;

(3) interpret the reconstructed protophonemes as a contemporaneous proto-phonology of a protolanguage;

(4) reconstruct the protolexicon by substituting the reconstructed protophonemes in the appropriate positions of the constructed cognate sets.[9]

This algorithm may not yield reliable and realistic historical reconstruction in cases where the historical divergence of lects is non-discrete. In such cases (as illustrated by Southworth above), the reason the method fails is that Steps 3 and 4 implicitly rely on essentialist notions of 'a language' and linguistic phylogeny. These notions lead to the false assumption that Steps 1 and 2 somehow guarantee the reconstructed protophonemes as contemporaneous constituents—a 'protophonology'—of a uniform and historical 'protolanguage'. However, as Geraghty concludes for Fiji in the quote above 'A feature found all over Fiji … may be a recent innovation'.

There are two ways of dealing with this problem of more recent changes reversing prior divisions between lects. Firstly, the problem can be explicitly acknowledged and systematically excluded from the scope of subgrouping. Koch (1996), as part of a more detailed algorithm for the Comparative Method (cf. footnote 9), outlines how such a systematic exclusion is normally practiced as part of the Comparative Method:

> 7. Where two or more languages have undergone the same change—*and this change must be ordered chronologically before other changes which are not shared by the languages in question*—posit (i) an intermediate protolanguage ancestral to just the languages in question (which are thus defined as a subgroup) and (ii) a single change that took place only once at some time intermediate between the protolanguage and the intermediate protolanguage. (Koch 1996:221, my emphasis)

The problem faced by this approach is *how to establish this chronology* given the findings from Geraghty (1983) outlined above? If the correspondences generated by reticulation of earlier phylogenetic divisions look identical to those generated by simple phylogenetic division, then how are reticulations to be identified in order to exclude them?

A further question for this approach is: *why is it necessary to exclude reticulation events from phylogenetic importance?* If such an exclusion is consistently applied then we must exclude all changes which are propagated through linguistically heterogenous speech

---

[8] Still broader definitions could be given, including the study of loan words, linguistic areas, etc. However, the breadth of definition given here is sufficient for the purposes of the present discussion.

[9] A similar, but more exhaustive algorithm of the steps involved in the CM is outlined by Ross and Durie (1996) and also by Koch (1996). Step (1) here corresponds to Ross and Durie's Steps (2)–(3) and Koch's Steps (1)–(3). Steps (2)–(3) here are collapsed in Ross and Durie's Step (4) as they make no distinction between (i) the reconstruction of protophonemes and (ii) their interpretation as a contemporaneous system. Koch's algorithm is considerably more nuanced than the one outlined in the text here, with Step (2) here corresponding to his Steps (4)–(6) and Step (3) corresponding to his Steps (7)–(9). Step (4) here corresponds to Ross and Durie's Step (7) and Koch's Step (10).

communities. The problem is, of course, that sociolinguistics has shown 'orderly heterogeneity' to be a natural and normal characteristic of languages and their speech communities (Labov 1966, 1994, 2001; Weinrich et al. 1968; L. Milroy 1987; J. Milroy 1992, 1997; Croft 2000).

The second valid way of dealing with the problem of common changes subsequent to differentiation of lects is: (a) abandon the essentialist assumption that 'a language' is structurally homogenous; (b) throw open the definition of phylogenetic 'subgrouping' to include those sets of lects defined by a Propagation Event (regardless of whether it occurred prior or subsequent to differentiation); (c) not assume that the reconstructed protophonemes automatically constitute a contemporaneous protophonology; and (d) develop further methods for sequencing the innovations (see §3.4.3 below), in order to satisfy the methodological desiderata (§3.1.1). Step (c) implies a narrower definition of the CM than the one given above.

CM (narrow definition):

(1) construct correspondences between phonemes in putative cognate sets;

(2) reconstruct protophonemes and subsequent changes which account for the correspondences as regular phonological reflexes.

Under this narrow definition the CM is stripped of as many assumptions about subgrouping as possible,[10] so as to prevent covert chronologies from sneaking into the reconstruction without warrant. When interpreting correspondences in a dialect continuum situation, the linguistic historian must keep in mind that *the shared features may have been more recently innovated than the divergent features*. Further criteria must be satisfied before the sequencing of shared and divergent features can be established.

3.3.3 Non-essentialist subgrouping

Before we move onto the methods for sequencing innovations, some discussion is in order regarding the notion of a phylogenetic subgroup. Traditionally, subgroups are defined on the basis of changes that precede any phylogenetic split between the lects in question. However, the exclusion of reticulation events from subgrouping importance is not a requirement of the sociohistorical theory outlined above. Within this framework, a subgroup is defined by the continuous transmission of linguistic material from a propagation defined language. The PDL is defined not by its structural homogeneity, but by a PE. Whether or not the range of this PE extended beyond the range of earlier PEs (creating overlapping isoglosses) is beside the point. This non-essentialist definition of phylogenesis does not insist on the Maxim 'once a subgroup, always a subgroup'.

Such a theoretical position is, of course, a considerable departure from the traditional family tree model of phylogenesis and subgrouping. However, even in evolutionary

---

[10] The hedging words are necessary here because even in Step (2) we cannot avoid some minimal considerations of subgrouping. A level of judgement is involved in assessing during the reconstruction of protophonemes how many tokens under a particular correspondence are sufficient to justify a protosegment rather than a set of loanwords. When a correspondence set is attested by only a few tokens, the best approach is to be conservative and put questionable sets to one side as a temporary measure. Their significance can be evaluated at a later stage in the process—after some progress in the reconstruction of chronology of changes and subgrouping relations—based on the correspondence sets attested by a greater number of tokens. I am grateful to Beth Evans for bringing this necessary qualification to my attention.

biology it is recognised that phylogenetic reticulation—reversal of prior phylogenetic division—occurs commonly in the evolution of both plants and parasites.[11] Plant species undergo hybridisation, and in addition, parasites undergo 'horizontal gene transfer' (as opposed to the 'vertical' gene transfer that occurs in reproduction of organisms). Morrison (2005:567) writes:

> Phylogenetic analysis has changed greatly in the past decade, including the more widespread appreciation of the idea that evolutionary histories are not always tree-like, and may, thus, be best represented as reticulated networks rather than as strictly dichotomous trees.

Huson and Bryant (2006:254P concur:

> The evolutionary history of a set of taxa is usually represented by a phylogenetic tree, and this model has greatly facilitated the discussion and testing of hypotheses. However, it is well known that more complex evolutionary scenarios are poorly described by such models. Further, even when evolution proceeds in a tree-like manner, analysis of the data may not be best served by using methods that enforce a tree structure but rather by a richer visualization of the data to evaluate its properties, at least as an essential first step. Thus, phylogenetic networks should be employed when reticulated events such as hybridization, horizontal gene transfer, recombination, or gene duplication and loss are believed to be involved, and, even in the absence of such events, phylogenetic networks have a useful role to play.

Though I do not believe it is necessary for historical linguistics to ape evolutionary biology, it may perhaps help persuade some readers of the viability of alternative phylogenetic models to learn that the necessity and usefulness of such models is recognised even within evolutionary biology.

Returning to the field of linguistics, Ross (1988) pioneered a categorisation of subgroups as either 'families' or 'linkages'. In Pawley and Ross (1995) the terminology was adjusted to 'innovation-defined' vs. 'innovation-linked' subgroups (subsequently adopted in Ross 1997; Thurgood 1999; Kirch and Green 2001 and Lynch, Ross and Crowley 2002). An innovation-defined subgroup is 'defined by shared innovations relative to a protolanguage' (Ross 1997). An innovation-linked subgroup, on the other hand, is a subset of lects characterised by non-coterminous innovations, with no innovation extending to the subgroup as a whole:

> whereby, say, languages A, B, C, and D reflect one innovation set, languages C, D, E, and F another set, languages D, E, F, and G another, and languages G, H and I yet another innovation set … When a group displays this kind of pattern of overlapping innovation sets, it is an 'innovation-linked subgroup'. In this case, we infer that its members are descended from an earlier dialect network. Innovations occurred in various dialects, each spreading into neighbouring dialects, but not across the whole network. The crucial point about an innovation-linked subgroup is that its innovations give us no evidence of an exclusively shared protolanguage. (Lynch et al. 2002:92 [see also the diagram on the same page])

Their definition of this category of subgroup has two components:

(1) the subset of lects in question is not derived from an exclusively shared protolanguage;

---

[11] See Mufwene (2001:152) for a discussion of similarities between linguistic history and the evolution of parasites.

(2) the precise phylogenetic relations between these lects cannot be determined beyond this level of detail. That is, the sequencing of the PEs is irrecoverable.

Thus, the categories of 'innovation-linked' vs. 'innovation-defined' are useful in contexts where (i) a dialect continuum stage is followed by discrete divergence of lects, and (ii) the sequencing of changes during the dialect continuum stage is unknown.

However, for studies such as the present one the aim is not just to label the whole dialect continuum as an 'innovation linked subgroup' and leave it at that.[12] Rather, the aim is to reconstruct as far as possible the details and chronologies of PEs that occurred during the dialect continuum's history. We can illustrate this goal using the abstract example quoted above from Lynch et al. The hypothetical data are characterised by four innovation sets:

Innovation set #1 {A,B,C,D}
Innovation set #2 {C,D,E,F}
Innovation set #3 {D,E,F,G}
Innovation set #4 {G,H,I}

There is no innovation common to all lects A-I, so the complete set of lects does not constitute an innovation-defined subgroup. However, each of the subsets #1-#4 is defined by a PE, and therefore each of these sets is of phylogenetic significance. Each of the sets #1-#4 defines a subgroup of lects *unless we insist on the theoretical impossibility or irrelevancy of reticulated phylogenetic divisions.* Under the theory presented here, reticulation is of phylogenetic significance and hence does not exclude sets #1-#4 from defining four distinct subgroups (cf. §3.3.2)

How would this dataset be approached within the sociohistorical framework proposed herein? First, the diagnostic value of the innovations would be considered, to ensure that the innovations are indeed diagnostic of PEs (the diagnostics are discussed below in §3.4.1). Second, we would use linguistic, textual, and (attested or hypothetical) sociohistorical criteria to sequence the changes if at all possible (cf. §3.4.3).[13] Third, the phylogenetic relations of the four subgroups would be modelled using an adjusted phylogenetic tree (cf. §3.4.4), rather than the traditional family tree.

If it turns out that the sequencing of innovations is not recoverable by the criteria outlined in §3.4.3, then saying that lects A-I constitute an innovation-linked subgroup is as much as we can do. More precise phylogenetic relations are only recoverable if the sequencing of innovations can be disambiguated. The methodology for this disambiguation is addressed in detail below.

---

[12] I am not saying that those who have employed the notion of an innovation-linked subgroup had any choice in the matter. Where it is impossible to recover the sequencing of innovations because of a lack of data, we have to cut our losses.

[13] Textual evidence is absent for Oceanic history, and sociohistorical evidence is also absent except for broad archaeological clues. As a result, the only criteria available as a basis for sequencing are linguistic (cf. §3.4.3.1) and hypothetical-sociohistorical (based largely on considerations of geographical plausibility of propagation, cf. §3.4.3.3).

3.4 Reconstructing Propagation Events in linguistic history

Recalling that it is propagation events (not structurally similar features) which define phylogeny, the linguistic historian must first verify that *the range of the innovative feature is the result of a propagation event*. There are two steps to this verification process (cf. Croft 2000:15–16):

(1) Ensuring that the feature is *innovative* by excluding (as far as possible) the possibility of retention from a more historically distant stage. In Croft's evolutionary terms, this amounts to excluding 'symplesiomorphies'—retained traits from an earlier parent population.

(2) Ensuring that the range of the feature is diagnostic of a single propagation event by excluding (as far as possible) the possibility of drift, or of independent and parallel propagation events.

The exclusion of specific changes as possibly the result of independent and parallel development must be principled and not ad hoc. The next section outlines a set of diagnostics which are intended to satisfy (as far as possible) the second step of historical verification.

3.4.1 Diagnostics for reconstructing Propagation Events

The diagnostics presented in this section have been developed with a particular sociolinguistic context in mind—the dialect continuum. Recall from §3.2.3 that novel variation of linguistic structure is the first mechanism of language change, and that this variation is conditioned by the linguistic structures already in place. However, instantiations of structurally similar innovations are only diagnostic of a propagation event if they are connected by interaction between speakers. Structural similarity alone is not what defines linguistic phylogeny. Accordingly, the diagnostics are not limited to structural criteria (Diagnostic #1), but also include sociolinguistic and sociohistorical criteria (Diagnostics #2 and #3) for reconstructing PEs in linguistic history.

3.4.1.1 First diagnostic: linguistic complexity of the innovation

The more complex the linguistic conditioning of an innovation, the less likely it is that the innovation was replicated and propagated independently. For this reason, phonological changes which are specified for particular morphological positions have greater diagnostic value than those which are phonologically general—unless the phonological conditioning is complex or unusual. Similarly, changes involving cognate inflectional morphology (paradigmatic relations between cognate forms) are unlikely to have been independently repeated and are thus of high diagnostic value (Nichols 1996). For example, the innovation of two protoverbal endings *-ɔw̃ and *-ɔ̃, with paradigmatic relations '1SG' vs. '1PL' respectively, is of high diagnostic value because it is highly unlikely that this precise combination of forms and functions was independently replicated in distinct PNs (cf. §6.4.1.2).

There are certain kinds of reductions in complexity which should be given a low diagnostic value based on this diagnostic. I have in mind here (a) the loss of a variant from a lect, and (b) the loss of an element in a construction. In the first case, if two lects both inherit two variants with the same function, then the regularisation of one of these variants

(and the consequent loss of the other) is of low diagnostic value. For example, the genitive case marker in proto Magadhan varied between *-kara* and *-kera*. The phasing out of one of these variants in favour of the other is not to be considered diagnostic of a PE because of the high possibility that such regularisation may have happened independently (see further §3.4.1.4).

In the second case, an element of a construction is *dropped out*, with a consequent change in the structure of the construction. Such constructional reduction is of lower diagnostic value than were an element to be *added* to the construction, because it is more difficult on linguistic grounds to exclude the possibility that the *dropping out* of an element occurred independently and in parallel. An example of this type of change is the deletion of nouns of multitude from plural pronominal constructions: *pronoun-GEN(-a) + noun of multitude* 'PL pronoun' > *pronoun-GEN(-a)* 'PL pronoun' (see §5.3.3). This change would be of greater diagnostic value for reconstructing a PE if new linguistic material had been *added* to the construction rather than *taken away.*

3.4.1.2 Second diagnostic: ecological distinctiveness of the innovation

In a dialect continuum, speakers of different varieties are rarely perfectly isolated from each another. Speakers of one variety often have contact with at least one politically official variety ('standard language'), as well as the varieties spoken by geographically and socially contiguous communities. The sociolinguistic context is thus characterised by diglossia and multilingualism (cf. Ferguson 2000).

I use the term 'linguistic ecology' or simply 'ecology' to refer to this sociolinguistic context, and 'ecological distinctiveness' to refer to whether an innovative feature is 'distinctive within the ecology'. That is, does the local innovation resemble features of the regionally superposed language? If so, can the innovation be construed as contact-induced through diglossia? If the answer to these questions is in the affirmative, then the change is not diagnostic of a PE.

For example, a diglossic relation to Hindi is part of the linguistic ecology of the Rangeli, Mahayespur and Kishanganj KRDS lects (three of the lects included in the reconstruction of Chapter 4) because these varieties are spoken either within, or in close proximity to, Bihar state where the official language and language of education is Hindi. In the linguistic data collected for these three KRDS lects, there is evidence of loanwords from Hindi as well as structural convergence with Hindi norms. Borrowing may be illustrated by [MI 8.], which introduces the non-inherited morpheme /sɛ/ 'INS'. The same form-function mapping is found in Hindi. Therefore this change is not ecologically distinctive, because it resembles the features of the regionally superposed language. Structural convergence can be illustrated by the unrounding of *ɔ > [ʌ] or [ə] in these same three KRDS lects (cf. §4.4.5). This unrounding process brings the pronunciation of shared NIA vocabulary in RL, MH and KS into closer conformity with Hindi norms. The similarity with Hindi of the novel variant phoneme /ʌ/ means that this change is not ecologically distinctive.

The implications of the non-distinctiveness of these changes in the ecology are as follows: the presence in RL, KS and MH of an unrounded pronunciation of *ɔ cannot diagnose a propagation event *connecting these three lects*. All three of the lects are in a diglossic relationship with Hindi, and so the novel variation could have entered each of the lects independently as a result of their similar linguistic ecologies. Similarly, the introduction of /sɛ/ 'INS' is not diagnostic of a propagation event connecting these three

lects because this morpheme was possibly borrowed independently in each of the sites as a result of diglossia.

The implications of ecologically distinctive changes are exactly opposite. They are not explained by diglossia, and are diagnostic of a propagation event. A case in point is the innovation of a singular and plural distinction in the secondary first person verbal endings (cf. §6.4.1). This change is attested across KRDS. However, the major superposed languages in the KRDS area—Bangla and Asamiya—lack this distinction. Therefore the change is ecologically distinctive: it cuts 'against the grain' of the pressure to conform to the linguistic structures of the superposed lects.

Changes that occur in a lect due to contact with a second, completely unrelated lect, are a further type of ecologically non-distinctive (and hence non-diagnostic) change. This type of change can again be illustrated from the history of KRDS. There are phonological features of KRDS which are structurally similar to corresponding features in eastern Bangladeshi varieties, namely the alveolar articulation (t͡s, d͡z,...) of the inherited laminal series (*t͡ʃ, *d͡ʒ,...; termed 'palatals' in IA studies). Chatterji (1926:79) hypothesises that this structural similarity is due to contact in both areas with Tibeto-Burman lects. If we accept this hypothesis, and it seems a good one, then this has implications for the diagnostic value of the alveolarisation of the inherited laminals (cf. §4.3.9). The alveolarised laminals in north and east Bengal are not diagnostic of a single PE, because this feature was possibly replicated independently due to Tibeto-Burman contact. This change does not, therefore, identify KRDS and eastern Bengali as a phylogenetic subgroup.

3.4.1.3 Third diagnostic: Sociohistorical plausibility of unified propagation across the range attested by the innovation

Face-to-face interaction between speakers leads to the propagation of novel variants (Trudgill 1986:39), and hence ranges of innovations may be expected to correlate with patterns of speaker interaction. A range which encompasses a geographically contiguous and socially connected 'zone' is plausibly explained as the result of propagation through face-to-face interaction. This is the scenario envisaged by the 'wave model' of language change. Sociohistorical plausibility (cf. §3.5) revolves around whether or not, for some stage in history, we have reason to reconstruct geographical contiguity and social connection which could account for the propagation of this innovation through face-to-face interaction. The sociohistorical argument may be based on attested social history, or in the absence of records, on hypothetical sociohistorical scenarios.

In addition to wave-like patterns of propagation, dialectological and sociolinguistic studies have found that the range of an innovation sometimes extends to larger population centres, skipping over the less densely populated areas in-between. This pattern of propagation is explained by the 'gravity' or 'hierarchical' model of language change (Trudgill 1974; Chambers and Trudgill 1998; Wolfram and Schilling-Estes 2003). A representation of the gravity model of change is given in Figure 3.1.

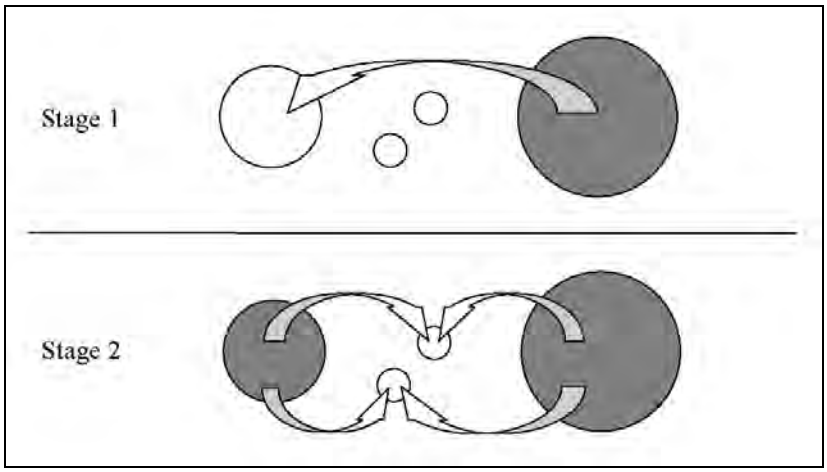


Figure 3.1: The gravity model of linguistic propagation

In the case of hierarchical patterns of propagation, the sociohistorical plausibility stems from considerations of speaker *density* as well as *distance*: the more dense the population of speakers, the greater the number of interaction events, and the more likely that propagation of a variant will occur through those interaction events. Consequently, a variant may be propagated between utterance pools of two large population centres before the propagation extends to the utterance pools of smaller intermediate population centres. Both the gravity and wave models of propagation should be kept in mind when making judgements about the sociohistorical plausibility of an innovation.

When the range of an innovation is neither geographically contiguous, nor explicable on the basis of hierarchical propagation between large population centres, then the onus is on the linguistic historian to provide a sociohistorical explanation that can account for the range as the outcome of propagation by speaker interaction. Without such a sociohistorical account, there may be little reason to conclude that the range is the result of a unified propagation event, and considerable reason to suspect independent replication followed by independent propagation events.

Thus, considerations of sociohistorical plausibility—by way of (attested or hypothetical) scenarios—have a significant role to play in distinguishing between single and disconnected propagation events.

3.4.1.4 Inherited variation with subsequent independent regularisation

There is a sub-type of non-diagnostic change which requires further mention in relation to dialect continua. In morphological systems we commonly find—whether by reconstruction or in historical texts—evidence for the historical presence of two forms serving the same general function, perhaps with subtle (and unreconstructible) differences in the grammatical, stylistic or sociolinguistic conditioning of each variant. In the descendant systems, one of these two forms may be regularised for the function in question, with the resultant loss of the other historical variant in that lect. The diagnostic value of this 'regularisation of inherited variation' is low.

As illustration of this phenomenon, consider the Genitive case forms in eastern Magadhan lects: *[-ɛr, -ɔr] < MIA [-*kera, -kara*] (the final vowel and medial *k* were lost by early NIA). The form *-ɛr is found in most contemporary Bangla dialects, as well as in the eight KRDS sites (see further §5.3.5). The *-ɔr form is found in Asamiya lects (to the

north-east) and in the BN site which is the KRDS lect most closely associated with Asamiya. So far the range of each variant is relatively contiguous, and the regularisation change might be proposed as diagnostic of a propagation event resulting from interaction between speakers. However, looking further afield, the /-ɔr/ form is also found in *Kharia Thar* to the south-west of KRDS (the opposite side to Assam), as well as in *Middle Bangla texts* alongside the /-ɛr/ form (Chatterji 1926:717).

In cases like this one, unless the regularisation of one variant has complex conditioning—thus arguing against parallel development by Diagnostic #1—the range of straightforward regularisation is not diagnostic of a unified propagation event because of the considerable possibility that regularisation occurred through independent and parallel propagation events. In the example above, neither of the two variants is regularised with complex conditioning—it is simply a matter of one being selected over the other as the marker of Genitive case. If a plausible argument can be made that the variability is inherited from an earlier common stage (either based on comparative reconstruction or historical texts), then the regularisation of this variability is not diagnostic of a PE because of the possibility that it occurred independently and in parallel. Consequently, the distribution of *-ɔr in Assam and in Kharia Thar is not diagnostic of a propagation event—it is more likely to be the result of independent and parallel regularisation of the *-ɔr variant. Likewise, the regularisation of the *-ɛr variant (across KRDS, Bangla lects, Mal Paharia, etc.) is not diagnostic of an interconnected propagation event.

The point is not that speaker interaction was not causal in the propagation of these changes at some local levels, but that there are *no linguistic criteria* for distinguishing which parts of the range were connected developments (PEs) and which were independent and in parallel. This class of changes, non-diagnostic of propagation events, is referred to throughout this study as *inheritance of variation with subsequent regularisation*.

3.4.2 Propagation Events (PEs) and Speech Community Events (SCEs)

We have seen above that plausible sociohistorical scenarios can play a role in diagnosing unified propagation events (§3.4.1.3). The causal relationship between social interaction and propagation of linguistic changes also means that, where there is a disjunction between the ranges of successive PEs, this disjunction may be evidence of *an intermediary change in the social structure of the speech community*. This method of using linguistic evidence to reconstruct changes in social structure has been developed by Ross (1997) with an eye to recovering aspects of the unrecorded social and cultural history of the Oceanic-speaking peoples.

Given the immediate context of the preceding section, a brief caveat too important to be relegated to a footnote needs to be added to this reconstruction principle. Propagation events *that are used to reconstruct changes in social structure* must be established by either Diagnostic #1, Diagnostic #2, or Diagnostic #3 based on *attested* sociohistorical scenarios only. Reconstruction of social history *cannot* be based on PEs which in turn are diagnosed primarily on the basis of *hypothetical* sociohistorical scenarios (a subcategory within Diagnostic #3). This restriction is required in order to avoid circularity of the following kind:

> Hypothetical sociohistorical scenario S_2 provides support for linguistic change X > Y to be considered the result of $PE_{X>Y}$—this in spite of no great linguistic complexity (Diagnostic #1) or ecological distinctiveness (Diagnostic #2).

> Disjunction between the range of $PE_{X>Y}$ and other historically antecedent PEs is used as evidence for a Speech Community Event (SCE), which derives scenario S_2 from a prior state, S_1.

This circular sociohistorical reconstruction is avoided: if PEs are diagnosed either by Diagnostics #1, #2, or #3 with reference to *attested* sociohistorical scenarios; or if PEs diagnosed by *hypothetical* sociohistorical scenarios are not admitted as evidence for reconstructing SCEs.

Ross' method of sociohistorical reconstruction is based upon the crucial distinction between a 'linguistic event (an innovation in a language)' and a 'speech community event (a change in the life of that community)' (ibid.:214). Ross holds that 'an innovation becomes part of the history of the language only when it spreads through the network to become a stable feature in the speech of a group of speakers' (ibid.:214-215). Ross' terms 'linguistic event' and 'linguistic innovation' have the same reference as the term 'propagation event' (PE) used in this study (defined in §3.2.1).

Speech Community Events include 'major changes in [social] network size and/or structure' (ibid.:215). These major reshapings of network structure may be reconstructed from disjunctions between PEs on two conditions. First, the sequencing (relative chronology) of the PEs must be known. For example, let us say that PE_1 was followed by PE_2, with a disjunction between the Propagation Networks (PN_1 and PN_2) diagnosed by the two changes. If the range of PN_2 is more *restricted* than PN_1, then this may be evidence for the creation of a communicative division in the Speech Community between the time of PE_1 and PE_2. If on the other hand PE_1 was followed by a more *expansive* PE_2—the range of PN_2 including and surpassing PN_1—then this may be evidence for a communicative re-integration in the Speech Community at some time after PE_1 and before PE_2.

The second condition for reconstructing SCEs—which explains the hedged use of 'may be' in the previous paragraph—is that the disjunction between PNs must be *more perspicuously explained by a change in SC structure rather than as the continuation of previously existing PNs within a single complex Speech Community*. To continue the hypothetical example above, it may be that the disjunction between PN_1 and PN_2 does not reflect a change in the SC structure, but just reflects co-existing networks within a complex SC. For the disjunction in PNs to be diagnostic of a SCE, the SCE must be more plausible than the possibility of co-existing complex social structures of speaker interaction.

To summarise the methodological principle outlined in this section:

(i) Given the causal relationship between patterns of social interaction and patterns of propagation of linguistic innovation;

(ii) If
   a. there is a disjunction between two PNs,
   b. and the sequencing of the PEs is reconstructible,
   c. and a sociohistorical change event is, in this case, more plausible than the co-existence of the PNs within a complex SC,

(iii) Then a SCE—consisting of either *division* or *integration* of communicative relations—is reconstructed.

3.4.3 Sequencing PEs

In §3.3 it was argued that the CM alone cannot guarantee the chronology, or contemporaneous status, of reconstructed changes. Further criteria are necessarily applied to the results of comparative reconstruction in order to sequence the reconstructed innovations. This section outlines three types of criteria—linguistic, textual and sociohistorical—which provide principled reasons for reconstructing the chronology of innovations. They are supplementary to the CM, but not optional for the reconstruction of linguistic history in a dialect continuum.

3.4.3.1 Linguistic seriation

Linguistic seriation is a term coined (as far as I am aware) by Anttila (1972:109) to refer to the sequencing of changes on the basis of linguistic criteria. Such criteria include the diachronic dependency of innovations, for instance:

> When the output of one change is the input of another, we can establish *relative chronology* between them.

Alternatively the criteria may not be *necessary* diachronic dependency, but *plausible* diachronic dependency—proposed on the basis of diachronic linguistic typology (see Koch 1996). From the mass of studies of various linguistic histories we learn that certain changes—whether phonological, morphological, semantic, or syntactic—commonly proceed incrementally, and in a certain order. For example, if the correspondences for lect A support reconstructing change *s > Ø, while the correspondences for lect B support instead *s > h, it is plausible to assume on linguistic criteria (that is, linguistic seriation) that *s > *h was common to both lects, with *h > Ø a linguistically natural incremental extension in lect A of the first change (see further Anttila 1989:§14.6). Anttila also uses linguistic seriation to sequence internally reconstructed changes (1989:§19.8–§19.10). Linguistic seriation can only be used to establish relative chronology of changes, and not absolute chronology.

3.4.3.2 Textual sequencing

Use of historical texts for linguistic reconstruction is not without its problems in Indo-Aryan (or elsewhere for that matter). Some comments have already been made regarding these problems in §1.1. Nonetheless, not all texts in the historical literature of Indo-Aryan are equally problematic. Those texts that aim to produce something closer to the vernacular of the time—usually prose rather than poetry—may, with some caution, be used to sequence innovative features.

In general the following principle should be followed when assigning chronology of innovations based on textual evidence: assume that the text is at least partially archaic and that the vernaculars of the time are more progressive in the use of innovative features than the attested written language. The implication of this interpretive principle is that the presence of an innovative feature in an historical text is good evidence that the feature had occurred in some lect at the time of writing.

The interpretation should, however, be asymmetric. If an innovative feature is absent from an historical text, this is not necessarily evidence that the innovation had *not* occurred in the vernaculars of the day. It is entirely plausible that the written norms lagged behind the spoken norms of the day.

Further problems associated with using historical documents to sequence linguistic changes are (1) establishing the chronology of the document in the first place—this is often done on the basis of references to historical events, as well as analysis of the style of script used—and (2) establishing the language in which the document was written. The latter problem is particularly notorious for documents written during early NIA. Assigning chronology to innovations based on documents whose ancestry and identity is controversial should be avoided. However, where the ancestry and identity is more accepted, textual evidence may contribute significantly to sequencing changes in linguistic history.

3.4.3.3 Sociohistorical linguistic seriation of innovations

A third set of criteria for sequencing PEs is sociohistorical in nature. As discussed in §3.4.2, a disjunction in range between different PEs may be the result of a change in SC structure. We have seen that for changes in SC structure to be reconstructible from PEs, the sequencing of the PEs must first be known. However, it is also possible to *reverse* the direction of sociohistorical reconstruction and use social history to sequence PEs. The shape of the historical argument is as follows:

(i) If PNs are reconstructed with a disjunction in their ranges,
   a. And a SCE is, on balance, more sociohistorically plausible than the co-existence of these PNs within a complex SC,
   b. And a particular directionality of SCE (i.e. either SC division or integration) is more plausible for sociohistorical reasons,

(ii) Then the plausible direction of the SCE supports a particular sequencing of PEs.

The sociohistorical plausibility for a particular direction of SCE may come from attested or hypothetical sociohistorical scenarios. Thus this method can be used both in areas with documented social history, and in areas without such documentation. This sociohistorical method of reconstruction is applied to the linguistic history of KRDS in Chapter 7, with good results.

A reminder regarding hypothetical sociohistorical scenarios: these can be used (a) to justify the diagnosis of a PE (§3.4.1.3), and as just stated, (b) to justify a particular sequencing of PEs. However, if hypothetical, unattested sociohistorical entities are used to establish both (a) diagnosis and (b) chronology of a PE, then the reconstruction is less historically reliable and more a matter of guesswork. The reconstruction becomes more historically reliable when either chronology or diagnosis of PEs is established by non-hypothetical criteria: either linguistic (Diagnostics #1 and #2, and linguistic seriation), textual sequencing, or *attested* social history.

The use of social history as a means of establishing the chronology of linguistic changes is not entirely new to historical linguistics. Inferences of this kind can be found in several studies but the methodology for developing these inferences has nowhere (that I know of) been made explicit and proceduralised. Geraghty (1983) cites Pawley and Sayaba (1971) as employing this approach:

> Combining their analysis of the linguistic situation with archaeological and topographical considerations, Pawley and Sayaba suggested that most of the favourable coastal regions of all the main islands of Fiji had been settled by 1,000 BC (Geraghty 1983:351).

In this case, the archaeological considerations attest ancient SCEs, and provide justification for *dating* those SCEs.

Topographical considerations, on the other hand, suggest hypothetical sociohistorical scenarios. These can be used to argue for a plausible direction of SCE:

> The coastal regions of Viti Levu and Vanua Levu [of Fiji], and the smaller islands, are all fairly readily accessible to each other by sea. The only really effective barriers to movement in Fiji are the mountain ranges … Once a sizeable proportion of the Viti Levu population moved inland up the main river valleys [as a result of decreasing reliance on maritime resources and/or overpopulation in coastal areas], the central mountain chain would have neatly divided the inland population into two. (Pawley and Sayaba 1971:433, cited in Geraghty 1983:351)

In the European context, data from cultural history (cultural centres, diocesan boundaries and the like) have been correlated with the ranges of linguistic innovations in Bloomfield (1935) and Lehmann (1992). Sociohistorical events support the dating of PEs correlating to the Benrath line as post-13th century changes (Lehmann 1992:127ff.); PEs which stop at the Ürdingen line are dated before AD 1789 (Bloomfield 1935:344–345).

In the course of this study, the following procedure for sequencing linguistic changes has been developed and tested:

I. Reconstruct the directionality of linguistic changes (e.g. by the CM).

II. Scrutinise in as much detail as possible the social and geographical ranges of the linguistic innovations established under Step I.

III. Apply the three diagnostics (linguistic complexity, ecological distinctiveness, and sociohistorical plausibility) to the innovations reconstructed under Step I to establish PEs in linguistic history.

IV. Investigate whether the chronology of any PEs that result from Step III can be established (a) by linguistic seriation involving *necessary* diachronic dependency or *plausible* diachronic dependency (cf. §3.4.3.1), or (b) by textual sequencing.

V. Consider (i) the possible permutations of SCEs (divisions and integrations) which would account for the disjunction in PNs, (ii) the relative sociohistorical plausibility of each possible permutation, and (iii) the relative sociohistorical plausibility of a SCE as against the co-existence of the PNs within a complex SC. Accordingly, reconstruct the chronology of PEs by selecting the most plausible sociohistorical explanation.

VI. Use the chronologies established by sociohistorical linguistic seriation (Step V), as well as linguistic seriation and textual sequencing (Step IV) to reconstruct an account of the linguistic history.

There is a parallel between the plausibility considerations involved in linguistic seriation and those involved in sociohistorical linguistic seriation of PEs. In linguistic seriation, the notion of plausibility is informed by ‘diachronic linguistic typology’ (cf. §3.4.3.1). Analogously, the notion of plausibility in sociohistorical linguistic seriation of PEs should be informed by a ‘sociohistorical linguistic typology’. Some steps towards such a typology are taken in §3.5.

3.4.4 Modelling PEs with phylogenetic tree diagrams

Based on the finding in Southworth (1958)—that structurally differentiated systems can undergo common innovations—Southworth (1964) tackles the problem of how to represent with a diagram this kind of phylogenetic history. In the terms of this study, the problem is how to depict both phylogenetic relations that result from SC integration and those that result from SC division. Southworth provides various diagrams as possible ways to synthesise the descriptive strengths of both the family tree and wave models of language change. His 'diachronic isogloss map' is given below as a representation of the phylogenetic relations of Panjabi, Hindi, Bengali, and Marathi. Each node in the tree represents a protolanguage. Diagonal and vertical lines show genetic transmission, and enclosed areas indicate ranges of propagation:

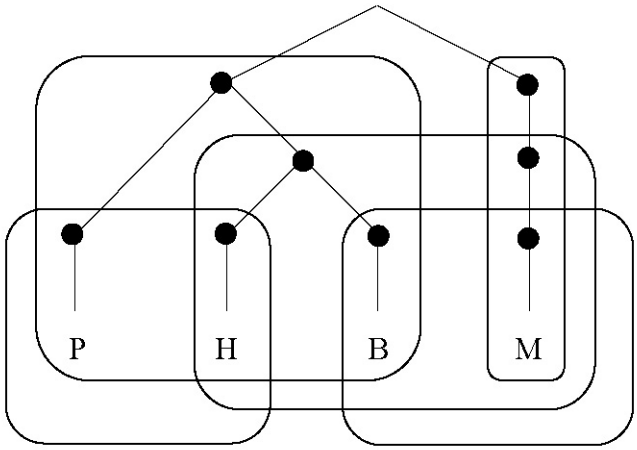


Figure 3.2: Southworth's 'diachronic isogloss map', illustrated for Panjabi, Hindi, Bengali and Marathi

This diagram—in particular the overlapping isogloss lines—is quite messy. While the messiness may reflect the historical reality, it may alternatively indicate a weakness in the method of representation. The most concerning characteristic of this model is that certain phylogenetic subgroups are marked by a node in the tree while other phylogenetic subgroups (as I use the term) are not. The node obscures the phylogenetic equality of, for example, subgroups *Panjabi-Hindi-Gengali and *Panjabi-Hindi.

Ross (1997, 1998) proposes a different reworking of the family tree diagram, by marking recombined phylogenetic relations that result from SC integration with a double horizontal line. In his overall model of language change these double lines also indicate a protolinkage which defines an innovation-linked subgroup. The convention of using double horizontal lines is adopted here, but in the context of a somewhat different model of linguistic phylogenesis.

Given that phylogeny is defined by PEs, these constitute the 'nodes' of the phylogenetic tree. However, rather than depict them as nodes per se, I will use the double horizontal line—the horizontal extension of the line symbolising the spatial extension of the propagation event. Thus for phylogenetic relations resulting from SC division and SC integration alike, the Propagation-Defined Language is marked with a horizontal line. The method of representation that results is illustrated below, using the language relations reconstructed by Southworth (1958, 1964):

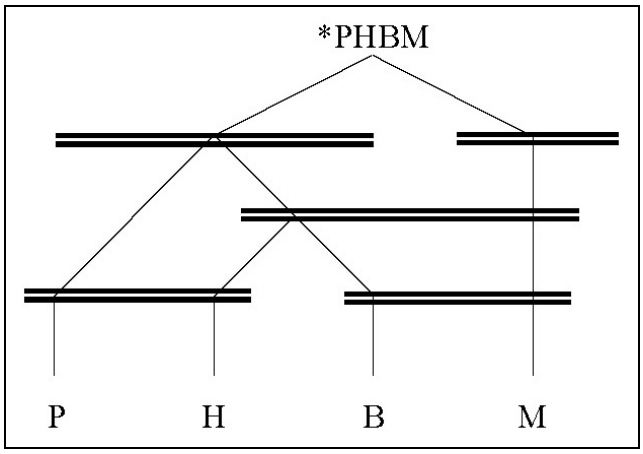


Figure 3.3: Adjusted phylogenetic tree diagram, illustrated for *Panjabi-Hindi-Bengali-Marathi (based on Southworth 1958, 1964)

This adjusted phylogenetic tree diagram depicts PEs occurring in PNs, and the phylogenetic relations between PDLs that result. The diagram is not, first and foremost, a model of SCEs. However, if the relative chronologies of the PEs are correctly depicted in Figure 3.3, then it becomes a relatively straightforward matter to infer the directionality of SCEs based on the PEs depicted in the diagram.[14] Figure 3.3 includes two phylogenetic subgroups which result from SC division: *Panjabi-Hindi-Gengali and *Marathi; and three phylogenetic subgroups which result from SC integration: *Hindi-Bengali-Marathi, *Bengali-Marathi and *Panjabi-Hindi.

Two further conventions are necessary in the application of this diagrammatic schema in the history of KRDS (see Chapter 7). Vertical or diagonal lines with an arrow head are used to show *established chronology*; lines without the arrow head are used to show that the chronology of PEs *has not been resolved*. Diglossic relations are shown with *broken* horizontal lines (cf. Figure 7.20), thus contrasting with the solid double horizontal lines of a PE. A broken *single* horizontal arrow is used to indicate the relation between the High and Low lects in diglossic relations. For example, in Figure 7.20, a broken single horizontal arrow connects RL and KS (two KRDS lects) with Hindi (the High language in that case).

3.5 Towards a sociohistorical linguistic typology

The sociohistorical linguistic seriation of changes is guided by notions of plausibility which will remain vague unless we attempt to construct a sociohistorical linguistic typology. The role of this typology in reconstruction is analogous to the role played by diachronic linguistic typology in linguistic seriation. The goals of sociohistorical linguistic typology should be to identify (a) sociohistorical criteria which are (probabilistically) diagnostic of a particular SCE, and (b) linguistic and sociohistorical features which (often) accompany each type of SCE and might guide the selection of a 'most plausible SCE'.

[14] I am not here claiming that the chronological relations in the diagram are the correct ones for the linguistic histories of Punjabi, Hindi, Bangla and Marathi, but simply showing how, if the chronology of PEs is known, the SCEs may be inferred from the diagram.

Ross (1997) is an important step in the direction of such a typology, and more such work is required—both in contexts with a recorded social history, as well as contexts without it. A thoroughly researched typology is well beyond the scope of this study. Nonetheless, the rest of this section outlines some thoughts in that direction, which should accordingly be taken as a first approximation (or, perhaps second, after Ross 1997) of a generalised sociohistorical linguistic typology.

A propagation event is diagnostic of two phenomena: communicative interaction and communicative isolation. Interaction enables propagation to succeed, isolation causes it to fail. Isolation leads to isogloss boundaries. Interaction, however, is not the sole factor that conditions the success or failure of propagation. Interaction must be accompanied by *common-identification* for a propagation to succeed. Sociolinguists (after LePage and Tabouret-Keller 1985) describe language use as involving 'an act of identity' whereby 'individual users of language strategically deploy varieties and variation to affiliate themselves with groups with which they may from time to time wish to be associated, or conversely, to be distinguished from groups with which they wish no such association' (Mendoza-Denton 2002:487). Keller's Maxims of communication include: 'talk in such a way that you are recognized as a member of the group' (1994:100). Mendoza-Denton notes a word of caution, that there exists a 'dissonance between automaticity and intentionality' in language use, which has as yet unexplored implications for speech events as acts of identity (ibid.:492). Nonetheless, the point stands that an utterance can have a social value assigned to it, and be used accordingly, even if the speaker is not conscious of the social value at the moment of utterance. Usage of linguistic variants which is automatic, but still identity-conditioned, is explained by the psychological theory of 'unconscious appraisals' taken up by Enfield (2003:9, following LeDoux 1998):

> we can, and continually do, make *unconscious appraisals* of the situations we find ourselves in (LeDoux 1998:64–65). Judgements and decisions in linguistic interaction can be made beyond our awareness, and thus may be beyond the reach of introspection.

It is thus theoretically sound, as well as empirically supported, that the propagation of variants in the usage of speakers results from sociolinguistic factors which in turn stem from considerations of speaker *interaction* and *identification*. Conversely, barriers to propagation are the result of:

(1) an inadequate (frequency or density of) *interaction* with speakers who have adopted the variant, and/or

(2) an inadequate sense of common *identity* with speakers who have adopted the variant.

The first type of barrier to propagation is any structure that hinders interaction between mutually identifying people, such as a geographical barrier or a (secondary) social barrier. For example, we might imagine a Bangladeshi Rajbanshi Hindu who wishes to interact with Indian Rajbanshi Hindus on the basis of their shared identity derived from centuries of tradition, but is unable to do so because of the relatively recent international border that now separates their communities. Similarly, the courses of sufficiently large and fast running rivers in some socio-cultural contexts lead to increased communicative isolation between communities on either side, regardless of their shared sense of historical identity.

The second kind of barrier to propagation is the presence of social divisions in a community. For example, in some areas of KRDS, the local Muslim community has undergone different innovations to the local Hindu community.

The goal of a sociohistorical linguistic typology is to produce general statements of the following abstract kind:

> Given sociohistorical conditions S_1, S_2, etc., we expect the outcome for the propagation of linguistic variants to be X, rather than Y.

In order to move towards statements of this kind, the foregoing discussion of sociolinguistics suggests we begin with the question:

> *What kinds of sociohistorical events might result in a change in the social structuring of (a) interaction patterns, and (b) relationships of identity?*

The first category of event, relating to changes in interaction, includes migrations, changes in marriage patterns, changes in a river's course, flooding, outbreak of disease in an area, new technology enabling communication and travel further from home, a bridge to cross a river, clearing of a forest area. Different events will be of greater or lesser relevance in different parts of the world, and there are certainly many other events that alter interaction patterns across the diversity of human cultures. Any of these events would be expected to have an effect upon potential propagation events by either removing or introducing new 'barriers' or 'bridges' between social networks.

The second category of event, related to changes in identity relationships, would include religious conversion, increased education, urbanisation, change in political situation, nationalisation, any successful social or political movement, or any other change in social identification. Each of these events may either introduce a new social identity, make obsolete an old identity, or change the relative salience of the multiple identities that feature in the lives of speakers.

Other questions that may guide research into the social conditioning of propagation will focus on the sociohistorical reasons for the extent (rather than the limits) of propagation. That is, the reliability of a sociohistorical reconstruction is strengthened by answering the following questions:

(1) Given an absence of the barriers listed above, what positive evidence is there outside of linguistic history for the presence of social continuity across the range of propagation? For example, what positive evidence apart from common linguistic innovations is there for interaction and identification between speakers of KRDS varieties in Cooch Behar and Rangpur?

(2) Secondly, if similar boundaries to those associated above with the limits of propagation are found *within* the geographical and social range of a propagation event, what sociohistorical explanation can be given for this apparent contradiction? For example, given that certain rivers have been significant in limiting PEs, why were other rivers not limiting factors?

By identifying the effects of sociohistorical events on propagation, answers to these questions would provide a sociohistorical linguistic typology. Such a typology would then inform and guide the sequencing of linguistic innovations by the procedure formulated in §3.4.3.3.

3.6 Conclusion

By way of concluding this chapter, let us return to the desiderata of §3.1 and briefly reflect on how this methodological framework satisfies those requirements.

1. Reconstruct change events in linguistic history; that is, linguistic innovations.

The change events in linguistic history are construed under this approach as involving (i) linguistic innovations, which are reliably reconstructed by a narrow definition of the CM; and (ii) the propagation events by which the innovation becomes established usage for a community of speakers.

2. Reconstruct the historical sequencing of these change events.

The framework articulated here provides three kinds of criteria for sequencing innovations and propagation events.

3. Reconstruct continuous transmission in linguistic history; that is, linguistic inheritance.

Continuous transmission of linguistic material is reconstructed by the narrow definition of the CM. Recurrent correspondences in cognate items are used to reconstruct protophonemes which represent a line of continuous transmission of words and morphemes (an etymology) through various stages in linguistic history. The contemporaneity of reconstructed protophonemes, as for protomorphemes and protowords, is not assumed on the face of it, but is justified by the reconstruction of the chronology of changes.

4. Reconstruct the time-depth for continuous transmission of a linguistic feature. This time-depth may be in terms of either 'absolute' time, or, more commonly, 'relative' to the time-depth of other features.

The time-depth problem is partially answered by the reconstruction of etymologies which are interrupted by sequenced changes. However, where an etymology is not affected by the reconstructed changes, the time-depth of inheritance remains ambiguous.

5. Reconstruct the historical relations between contemporary lects. Such relations are based on continuously transmitted material that testifies to a common change event.

Historical relations are defined within this framework by PEs, reconstructed on the basis of diagnostic innovations. The sequencing of PEs is displayed graphically by an adjusted tree diagram which models the phylogenetic relations between compared lects.

6. Consequently reconstruct the linguistic material (words, morphemes, phonemes) present at every stage in linguistic history.

This statement represents the ideal end-point—the summit of Everest, which the linguistic historian attempts to conquer. In the case of KRDS the framework developed here has been highly successful in advancing some way up the mountain. Gaps of understanding remain, especially concerning the time-depth of some inherited features. Nevertheless, this framework has arguably advanced the cause of reconstructing linguistic history in a dialect continuum, as is demonstrated by the reconstruction that follows for KRDS.

4 *Phonological reconstruction*

4.1 Introduction

This chapter begins the work of reconstructing the linguistic history of KRDS. Readers of this chapter should keep in mind that the common ancestor for the KRDS lects is reconstructed in §7.3.1 as 'proto Kamta'. This term is chosen for historical reasons explained in that section, and is distinct from the present-day political usage of the same term (cf. §1.2.2).

The method employed for phonological reconstruction is the Comparative Method which involves examining phonological correspondences across suspected cognates and reconstructing protophonemes and subsequent changes.[1] Three diagnostics—linguistic complexity, ecological distinctiveness, and sociohistorical plausibility—are applied to each reconstructed innovation in order to diagnose propagation events (and exclude parallel events not connected by interaction). These diagnosed PEs are the defining moments in KRDS's phylogenetic history and are the basis for reconstructing historical Speech Community Events in Chapter 7.[2]

As discussed in Chapter 3, reconstructing the sequencing of innovations is problematic in a dialect continuum because of the normalcy of non-discrete divisions between lects. The boundaries of community interaction are often diachronically unstable, and the result is overlapping and non-nested isoglosses. Moreover, the sequencing of these changes may be ambiguous based on solely *linguistic* principles of seriation. This study employs three types of argument for sequencing innovations: linguistic, textual, and sociohistorical.[3] Of these, the first two lines of argument are pursued in this chapter (as well as in Chapters 5–6 for innovations in morphology) to establish some relative and absolute chronology of innovations. Where more than one account is possible of the change events that produced the attested correspondences, the task of this chapter is to determine if there are linguistic (or textual) criteria that render one interpretation of the change events more plausible than other interpretations. The third line of argument for establishing chronology—by sociohistorical criteria—is applied separately in Chapter 7 to round out the reconstruction of linguistic history.

---

1 Compare §3.3 This method is called 'controlled reconstruction' in Pattanayak (1966) and Maniruzzaman (1977).

2 Compare §3.4.1 for justification of the diagnostics, and §3.2.2 for discussion of the phylogenetic model.

3 Compare §3.4.3.

The data which inform the phonological reconstruction may be found as a comparative wordlist in Appendix A of Toulmin (2006). Phonological correspondences have been tabulated using the WordCorr program (cf. www.wordcorr.org). The full set of correspondences for the KRDS reconstruction is listed in Appendix B of Toulmin (2006), with generalised correspondences presented in tables throughout this chapter.

The correspondence sets do not include data for Standard Colloquial Bangla (SCB), Asamiya (SCA) or other NIA lects. Nonetheless, the phonological innovations reconstructed here for KRDS can (with a little care) be compared and contrasted with innovations reconstructed in other historical Indo-Aryan studies whose method is either comparative (Southworth 1958; Pattanayak 1966; Maniruzzaman 1977) or etymological (e.g. Chatterji 1926; Kakati 1962; Turner 1962–66; D.N. Das 1990).[4]

4.2 Synchronic overview of systems and processes

While the purpose of this chapter is to present a thorough diachronic phonological analysis, an interpretation of the synchronic systems and processes underlies the transcription of data and consequently the reconstruction. This section sketches the phoneme systems of the eight lects included in the historical reconstruction. The synchronic analysis is based largely on the wordlist data in Appendix A of Toulmin (2006) as well as on synchronic studies referenced below. An exhaustive phonological study of 'central' KRDS—the varieties spoken in the districts of Cooch-Behar and (greater) Rangpur—remains to be undertaken or published.

The most thorough phonological analysis of a KRDS variety to date is Wilde (2008: Ch.2), which describes 'the phonemic elements of three of the Rājbanshi dialects spoken in the Jhāpā district (Korobāri, Lakhanpur and Ghailāḍubbā-Lakharigaddi)'. Geographically two of these locations are quite close to Mahayespur, eastern Jhapa, which features in this reconstruction. Toulmin (2002a) includes some phonological analysis of the Mahayespur variety, and Toulmin (in press) describes in some detail the Bhatibari-Tufanganj variety of West Bengal. D.N. Das (1990) gives a phonological sketch of the Indo-Aryan lects of greater Goalpara and Kamrup regions of west Assam (thus including the Bongaigaon lect which also features in this reconstruction). In addition to these descriptions of KRDS, there are good phonological analyses available for neighbouring NIA lects (Ferguson and Chowdury [1960] for Bangla, G.C. Goswami [1966] for standard Asamiya, U. Goswami [1970] for western Asamiya).

KRDS varieties have six or seven segmental vowel phonemes—the difference hinging on the inherited *ɛ, and its split into /e/ and /ɛ/ in several lects. The situation is complicated by borrowing. KS and RL both have six-vowel systems /i, ɛ, a, [ə; ʌ], o, u/.[5] MH appears to have a seven-vowel system—potentially the result of borrowing Hindi and Persian

---

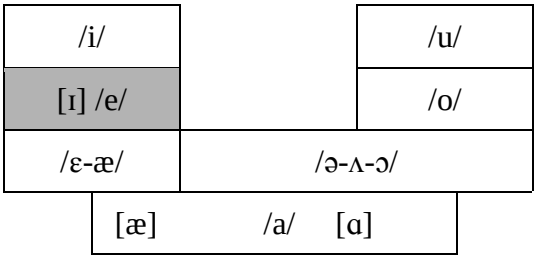
[4] By an 'etymological' method I mean that the study involves reconstruction of linguistic history by an uncontrolled comparison between putative cognates in Sanskrit and the lect(s) in question. The reconstruction is uncontrolled in that the regularity or otherwise of proposed sound changes is not a central or guiding concern. Despite the title of Das's thesis (1990): *The dialects of Goalpara and Kamrup: a comparative analysis*, the method employed is not the traditional Comparative Method of historical linguistics, but the 'etymological method' described here.

[5] /a/ is phonetically [ɐ] in all that follows, though Wilde considers it to be [æ]. /ə/ indicates a phonetically mid (in terms of height), central (in terms of backness) and unrounded vowel; /ʌ/ is phonetically unrounded, back and slightly lower than /ə/; [ə; ʌ] indicates synchronic variation.

words with a retained [e] in contrast with the inherited *ɛ, for example /ek/ ‘one’, /ɛk/ ‘to her/him/it’. But more rigorous analysis of MH’s phonology may find the situation to be considerably more complex than this. Describing the Rajbanshi lects west of MH, Wilde (2008:17–19) exposes a complicated allophonic distinction between /i/, /ɛ/ and [ɪ] in some Rajbanshi lects of Nepal. It should be noted that, from a diachronic perspective, the distinction described by Wilde between /ɛ/ and [ɪ] is almost certainly not the same as that referred to below concerning /ɛ/ vs. /e/.

The TH, SH, RP, and BH lects all have seven-vowel systems: /i, e, ɛ, a, ɔ, o, u/; in RP the corresponding vowel for /ɛ/ is lower and thus transcribed as /æ/. Note that the rounded vowel /ɔ/ of central and eastern KRDS (TH, SH, RP, BH, BN) corresponds with unrounded /ʌ/ of MH, and [ʌ; ə] of RL and KS. The unrounding in western KRDS is a result of Hindi influence (cf. §4.4.5), and found also in the Bihari lects. The status of the [e] vs. [ɛ] distinction has not been conclusively established for BN in this study, though it is assumed to be phonemic following D.N. Das (1990). These similarities and differences in vowel systems are summarised by Table 4.1.

Table 4.1: Amalgamated summary of present day KRDS vowel systems



Phonemes are enclosed by forward slash markers, and allophones in square brackets. The phonemic ‘slot’ which distinguishes the six and seven vowel systems is shaded.

The sound [ɑ] is found in BN, BH and RP, and assumed to constitute an allophone of /a/—though the conditioning is not yet understood. A preceding high vowel triggers the slight raising and fronting of /a/ → [æ] in BH and RP. In RP the merger with /æ/ is complete, while in BH it may remain distinct from the corresponding vowel /ɛ/ for some speakers in some contexts.[6]

The phonemic status of /e/ differs among the lects. Most instances of [e] can be described as allophones of /ɛ/ by regressive vowel harmony. However, loan words (mainly Persian and Hindi) have introduced minimal or near-minimal pairs with the contrasting sounds. All these phonological processes are examined in more depth as part of the diachronic reconstruction.

Turning to the KRDS consonant systems, seven of the eight lects examined here follow the typical Indo-Aryan pattern—dominated by stops, distinguished at five points of articulation: bilabial, dental, apical postalveolar (traditionally ‘retroflex’ or ‘cerebral’ in IA studies), laminal postalveolar (traditionally ‘palatal’ in IA studies), and velar. This has remained the Indo-Aryan pattern from Old IA, through Middle IA (despite many phonological changes) into New IA.

---

[6] The degree of harmonic raising of /a/ is socially conditioned in Cooch Behar district: in uneducated speech, /a/ merges with /ɛ/ after a high vowel; in more educated or ‘town-style’ speech the two phonemes do not merge in this environment.

Table 4.2: Amalgamated summary of present day KRDS consonant systems

Moving (active) articulator:			Labial (lower lip)	Apical (tongue tip)			Laminal (tongue blade)	Dorsal (back of tongue)	Laryngeal (vocal folds)
Articulatory target region (passive articulator):			Labial	Dental	Alveolar	Post-alveolar	Alveolar,[a] Post-alveolar	Velar	Glottal
Oral stops and affricates	-vc.	-asp.	/p/ [pʲ pʷ]	/t̪/ [t̪ʲ t̪ʷ]		/t̠/ [t̠ʲ t̠ʷ]	/ts-ʧ/ [tsʲ]	/k/ [kʲ kʷ]	
	-vc.	+asp.	/pʰ/	/t̪ʰ/		/t̠ʰ/	/tsʰ-ʧʰ/	/kʰ/ [kʰʲ]	
	+vc.	-asp.	/b/ [bʲ bʷ]	/d̪/ [d̪ʲ d̪ʷ]		/d̠/ [d̠ʲ d̠ʷ]	/ʣ-ʤ/ [ʣʲ ʣʷ]	/g/ [gʲ gʷ]	
	+vc.	+asp.	/bʱ/ [pʱ]	/d̪ʱ/ [t̪ʱ]		/d̠ʱ/ [t̠ʱ]	/ʣʱ-ʤʱ/	/gʱ/ [kʱ]	
Nasal stops		-asp.	/m/		/n/ [nʲ nʷ]			/ŋ/ [ŋʲ ŋʷ]	
		+asp.	/mʱ/		/nʱ/			/ŋʱ/	
Fricatives	-vc.		[ɸ ɸʲ ɸʷ]				/s-ʃ/ [s ʃʲ ʃʷ]	[x]	
	+vc.	-asp.	[β]				[z zʲ zʷ]	[ɣ]	
	+vc.	+asp.	[βʱ]				[zʱ]	[ɣʱ]	
Rhotics		-asp.			/r/ [rʲ rʷ]	/ɻ-ɾ/			
		+asp.			/rʱ-ɾʱ-ɻʱ/				
Laterals		-asp.			/l/ [lʲ lʷ]				
		+asp.			/lʱ/				
Approximants			/w/ [wʲ]			/j/			/h/

[a] The passive articulator for this set spans alveolar and post-alveolar.

The BN lect stands apart from the other seven lects because it lacks the distinction of apico-dental versus apico-postalveolar (‘dental’ versus ‘retroflex’). In BN, as in SCA, these two series have merged into a new apico-alveolar series. This restructuring of the BN consonant system seems to be recent given that it is unquestionably present in the data collected for this study, but was not described by D.N. Das (1990), see further §4.3.6.

Aspirated counterparts to modally voiced (unaspirated) nasals, laterals and rhotics are found in KS, RL, MH and TH (see further below regarding the definition of ‘aspiration’).

As in the presentation of vowels above, Table 4.2 has consonant phonemes enclosed by forward slash markers, and allophones by square brackets. Special abbreviations used in this table (due to space restrictions) are: vc. for ‘voice’, and asp. for ‘aspiration’.

The recurrent allophonic processes found amongst KRDS lects are: palatalisation and labialisation of stops under certain conditions (cf. §4.3.3); postvocalic spirantisation or fricativisation of stops and affricates (cf. §4.3.5); and devoicing of initial voiced stops—this last change more regular in BN, and present but more variable in BH (full analysis in §4.3.1).

The point of articulation for the affricates differs across the lects. In KS, RL and MH they are articulated closer to the alveolar ridge than in TH which is more postalveolar (at least for the speakers interviewed for this study). In SH, RP and BH the series is articulated on the alveolar ridge itself, while the BN system lacks affricates altogether—making it once more the phonological odd-one-out (cf. §4.3.9 for the historical changes involved).

The inherited sibilant has a postalveolar articulation in TH, SH, RP, and BH. In KS, RL, MH and BN the articulation is closer to the alveolar ridge.

The BN lect possesses an alveolar series of stops, but no apico-dental and apico-postalveolar series of stops. This phonological system bears close resemblance to that of Asamiya—distinguishing stops at only three places of articulation, and lacking affricates.

For voiced and aspirated consonants (e.g. /g^{ɦ}/) the aspiration is breathy voiced and transcribed by /^{ɦ}/. In contrast, the aspiration of voiceless aspirated consonants (e.g. /k^{h}/) is transcribed by /^{h}/. The definition of ‘aspiration’ used in this study follows Ladefoged and Maddieson:

> aspiration is a period after the release of a stricture and before the start of regular voicing (or the start of another segment, or the completion of an utterance) in which the vocal folds are markedly further apart than they are in modally voiced sounds. This definition would allow for voiceless aspirated and breathy voiced aspirated sounds to be grouped together (1996:70)

The aspirated continuants (nasals, laterals and fricatives) are phonetically characterised by the presence of breathy voicing: /m^{ɦ}/=[m̤]; /l^{ɦ}/=[l̤]. Close phonetic study of aspirated nasals and laterals in Hindi shows that breathy voicing begins after a brief initial period of modal voicing (cf. ibid.:107–108, 201–102). The coordination of modal and breathy voicing in stops (again in Hindi) is basically the same (cf. ibid.:57ff.). In this study, stops and continuants which are characterised by the coordination of modal and breathy voicing are alike termed ‘voiced aspirated’ consonants.

The phoneme /h/ is often classed as a fricative in IA studies. However, modern study of phonetics and phonology supports a different classification. Ladefoged (1971) describes the sounds [h] and [ɦ] (of which KRDS /h/ is the latter) as *voiceless or breathy voiced counterparts of the vowels that follow them*. More recently, Ladefoged and Maddieson revised this description:

> as Keating (1988) has shown, the shape of the vocal tract during **h** or **ɦ** is often simply that of the surrounding sounds. In saying the word *ahead*, for example, there is usually a breathy voiced **ɦ** during which the formants are moving from those associated with **ə** to those associated with **ɛ**. Accordingly, in such cases it is more appropriate to regard **h** and **ɦ** as *segments that have only a laryngeal specification, and are unmarked for all other features.* (Ladefoged and Maddieson 1996:325–326, emphasis added)

As a consonant whose characteristics are determined by the surrounding vowels, /h/ is more appropriately classed as an *approximant*, rather than a fricative. In KRDS, as in Indo-Aryan more generally, the /h/ is 'voiced', meaning that the laryngeal specification is for breathy voicing. All further phonetic features are determined by the adjacent sounds.

The postaveolar series of stops (usually referred to in IA studies as 'retroflex') has an apical active articulator in KRDS (as in Hindi) rather than the 'sub-apical' (or perhaps rather sub-laminal) articulation of the Dravidian languages (cf. Ladefoged and Maddieson ibid.:26ff.). Ladefoged and Maddieson use the IPA symbols /t̻, d̻/ to denote sub-apical retroflexes; and the non-IPA symbols /ṭ, ḍ/ for apical 'retroflexes', that is postalveolars. The transcription in this study adheres instead to the IPA conventions /ʈ, ɖ/ for 'retroflex' phones (taken as including apical postalveolars). However, note that this usage fails to signal the articulatory differences between the stops in KRDS and Tamil (for example).

4.3 Comparative reconstruction of KRDS consonants

The consonant system reconstructed in this study for proto Kamta conforms to the broad Indo-Aryan type. The protophonemes and their generalised reflexes in each of the eight KRDS test lects are presented in Tables 4.3 through 4.14.[7] (Note that these reconstructed phonemes are only considered to constitute a contemporaneous protophoneme system as a result of the total reconstruction of chronology of changes in this study. The necessity of reconstructing the sequencing of changes before hypothesising a contemporaneous phonological system is a point made in §3.3)

The correspondences are categorised with respect to three phonologically significant environments: word-initial, intervocalic and word-final. Where further categorisation of intervocalic conditions is required (e.g. [i_a]) the more limiting condition is given in a note below the table. Other conventions used are: Ø 'deletion'; - 'data missing but expected'; ˜ 'nasalisation of preceding vowel'; blank cell 'phoneme does not occur in this position'; semicolon 'two synchronic variants'; new line within the same cell 'different correspondences in different words' (with conditioning or degree of variation noted below the table). In the notes below the tables, V indicates 'any vowel', N 'any nasal consonant', C' any consonant', # ' a word boundary'.

---

[7] The full array of correspondence sets yielded by the phonological reconstruction is found in Appendix B of Toulmin (2006).

Table 4.3: Summary of inherited consonants and their reflexes (initial position)

*		*b	*b^{ɦ}	*p	*p^{h}	*d̤	*d̤^{ɦ}	*t̤	*t̤^{h}
	KS	b	b^{ɦ}	p	p^{h}	d̤	d̤^{ɦ}	t̤	t̤^{h}
	RL	b	b^{ɦ}	p	p^{h}	d̤	d̤^{ɦ}	t̤	t̤^{h}
	MH	b	b^{ɦ}	p	p^{h}	d̤	d̤^{ɦ}	t̤	t̤^{h}
#_	TH	b	b^{ɦ}	p	p^{h}	d̤	d̤^{ɦ}	t̤	t̤^{h}
	SH	b	b^{ɦ}	p	p^{h}	d̤	d̤^{ɦ}	t̤	t̤^{h}
	RP	b	b^{ɦ}	p	p^{h}	d̤	d̤^{ɦ}	t̤	t̤^{h}
	BH	b b;p [a]	b^{ɦ} b^{ɦ};p^{ɦ} [b]	p	p^{h}	d̤	d̤^{ɦ} d̤^{ɦ};t̤^{ɦ} [b]	t̤	t̤^{h}
	BN	b;p [c]	p^{ɦ}	p	p^{h}	d;t [c]	t^{ɦ}	t	–

[a] Devoicing is variable and **non-persistent** (found in only one etymon in the data).
[b] Devoicing is variable and **persistent** in a **minority** of etyma.
[c] Devoicing is variable and **persistent** in the **majority** of etyma.

Table 4.4: Summary of inherited consonants and their reflexes (initial position cont.)

*		*ɖ	*ɖ^{ɦ}	*ʈ	*ʈ^{h}	*ʤ	*ʤ^{ɦ}	ʧ	ʧ^{h}
	KS	ɖ	ɖ^{ɦ}	ʈ	ʈ^{h}	ʤ	ʤ^{ɦ}	ʧ	ʧ^{h}
	RL	ɖ	ɖ^{ɦ}	ʈ	ʈ^{h}	ʤ	ʤ^{ɦ}	ʧ	ʧ^{h}
	MH	ɖ	ɖ^{ɦ}	ʈ	ʈ^{h}	ʤ	ʤ^{ɦ}	ʧ	ʧ^{h}
#_	TH	ɖ	ɖ^{ɦ}	ʈ	ʈ^{h}	ʤ	ʤ^{ɦ}	ʧ	ʧ^{h}
	SH	ɖ	ɖ^{ɦ}	ʈ	ʈ^{h}	z	z^{ɦ}	ts	ts^{h};s
	RP	ɖ	ɖ^{ɦ}	ʈ	ʈ^{h}	ʣ	ʣ^{ɦ}	ts;s	ts^{h};s
	BH	ɖ	ɖ^{ɦ} ɖ^{ɦ};ʈ^{ɦ} [a]	ʈ	ʈ^{h}	ʣ	ʣ^{ɦ} ʣ^{ɦ};ts^{ɦ} [a]	ts	ts^{h}
	BN	d;t [b]	t^{ɦ}	t	t^{h}	z;s [b]	s^{ɦ} [c]	s	s

[a] Devoicing is variable and **persistent** in a **minority** of etyma.
[b] Devoicing is variable and **persistent** in the **majority** of etyma.
[c] phonetically [s̤]

Table 4.5: Summary of inherited consonants and their reflexes (initial position cont.)

*		*g	*g^{ɦ}	*k	*k^{h}	*m	*m^{ɦ}	*n	*n^{ɦ}	*ɳ	*ŋ
	KS	g	g^{ɦ}	k	k^{h}	m		n l [c]			
	RL	g	g^{ɦ}	k	k^{h}	m		n l [c]			
#_	MH	g	g^{ɦ}	k	k^{h}	m		n l [c]			
	TH	g	g^{ɦ}	k	k^{h}	m		n			
	SH	g	g^{ɦ}	k	k^{h}	m		n			
	RP	g	g^{ɦ}	k	k^{h}	m		n			
	BH	g	g^{ɦ} g^{ɦ};k^{ɦ} [a]	k	k^{h}	m		n			
	BN	g;k [b]	k^{ɦ}	k	k^{h}	m		n			

[a] Devoicing is variable and **persistent** in a **minority** of etyma.

[b] Devoicing is variable and **persistent** in the **majority** of etyma.

[c] /l/ for initial *n in KS, RL and MH is a minority correspondence, with conditioning irregular. The irregularity is (in some way) residual from MIA l/n variation – whether directly inherited in KRDS or through later interaction with other lects is unclear (see analysis of (PI 14.) *l > n below Table 4.25).

Table 4.6: Summary of inherited consonants and their reflexes (initial position cont.)

*		*l	*ɭ	*l^{ɦ}	*r	* ʃ	*h	*w	*j
	KS	l			r n [a]	s	h		
	RL	l			r n [a]	s	h		
#_	MH	l			r n [a]	s	h		
	TH	l;n			r;Ø	ʃ	h		
	SH	n			r;Ø	ʃ	h		
	RP	n			Ø	ʃ	h		
	BH	n			r;n	ʃ	h		
	BN	l			r	h s	h		

[a] #_VNC, e.g. *rand̪^{ɦ}- ‘cooks’.

Table 4.7: Summary of inherited consonants and their reflexes (intervocalic position)

*		*b	*bʱ	*p	*pʰ	*d̪	*d̪ʱ	*t̪	*t̪ʰ
	KS	w	β	p	ɸ	d̪	d̪ʱ	t̪	t̪ʰ
	RL	w	βʱ [a]	p	ɸ	d̪	d̪ʱ	t̪	t̪ʰ
	MH	β	βʱ	p	ɸ	d̪	d̪ʱ	t̪	t̪ʰ
	TH	b;β	bʱ βʱ	p	–	d̪	d̪ʱ	t̪	t̪ʰ
V_V	SH	β	β	p;ɸ pʷ [b]	–	d̪	d̪	t̪ t̪ʷ [c]	t̪ʰ
	RP	b bʲ [d]	– bʲ [d]	p;ɸ pʲ [b]	– ɸʲ [c]	d̪ d̪ʲ [b]	d̪ d̪ʲ [b]	t̪ t̪ʲ [b]	t̪ – [b]
	BH	β βʲ [d]	β;b bʲ [d]	p pʲ [d] pʲ;pʷ [c]	–	d̪ d̪ʲ [b]	d̪ d̪ʲ [b]	t̪ t̪ʲ [b]	t̪ – [b]
	BN	b;w	βʱ	p	–	d	dʱ	t	tʰ

[a] Throughout this table βʱ is phonetically [β̤].
[b] [i_a] and [u_a].
[c] [u_a] (Note: The data for some phonemes is position [i_a] is absent for some lects. Thus, further data collection is needed in some cases, and may show that conditioning enviornment is more general, i.e. 'after a high vowel, before /a/')
[d] [i_a] (if not given, the environment [u_a] may be missing from the data).

Table 4.8: Summary of inherited consonants and their reflexes (intervocalic cont.)

*		*ɖ	*ɖʱ	*ʈ	*ʈʰ	*ʤ	*ʤʱ	ʧ	ʧʰ
	KS	r	rʱ	ʈ	ʈʰ	ʤ	–	ʧ	ʧʰ
	RL	ɽ;r	ɽʱ;rʱ	ʈ	ʈʰ	z	–	ʧ	ʧʰ;s
	MH	ɽ;r	ɖʱ;rʱ	ʈ	ʈʰ	z	–	ʧ	ʧʰ
	TH	r	rʱ	ʈ	ʈʰ	ʤ	ʤʱ	ʧ	ʧʰ
V_V	SH	ɽ;r ɽʷ [a]	ɽ	ʈ	ʈʰ	z zʲ [b]	z	s	s
	RP	ɽ;r rʲ [c]	ɽ;r rʲ [c]	ʈ ʈʲ [b]	ʈ ʈʲ [c]	ʣ ʣʲ [c]	ʣ	ʦ;s ʦ;sʲ [c]	s sʲ [b]
	BH	ɽ;r rʲ [c]	ɽ;r rʲ [c]	ʈ ʈʲ [b]	ʈ ʈʲ [b]	ʣ ʣʲ [c]	ʣ	ʦ ʦʲ [c]	s sʲ [b]
	BN	r	r	t	tʰ	z	z	s	s

[a] [u_a] (if not given, the environment [i_a] may be missing from the data).
[b] [i_a] (if not given, the environment [u_a] may be missing from the data).
[c] [i_a] and [u_a].

Table 4.9: Summary of inherited consonants and their reflexes (intervocalic cont.)

*		*g	*g^{ɦ}	*k	*k^{h}	*m	*m^{ɦ}	*n	*n^{ɦ}	*ɳ	*ŋ
	KS	g	g^{ɦ}	k	k^{h};x	m	m^{ɦ}	n	n^{ɦ}	n	ŋ
	RL	g	g^{ɦ}	k	k^{h}	m	m^{ɦ}	n	n^{ɦ}	n	ŋ
	MH	g	g^{ɦ}	k	x	m	m^{ɦ}	n	n^{ɦ}	n	ŋ
	TH	g; ɣ	g^{ɦ}	k	x	m	m^{ɦ}	n	–	n	ŋ
V_V	SH	ɣ	ɣ^{ɦ}	x k$^{w\ a}$ k$^{j\ b}$	x -^{c}	m m$^{w\ a}$	m	–	–	n –^{c}	ŋ
	RP	g -^{c}	g -^{c}	k k$^{j\ c}$	k k$^{j\ c}$	m m$^{j\ b}$	m	n n$^{j\ c}$	n –^{c}	n n$^{j\ c}$	n –^{c}
	BH	g;ɣ –^{c}	g;ɣ –^{c}	k k$^{j\ c}$	x k$^{hj\ c}$	m m$^{j\ b}$	m	n n$^{j\ c}$	n –^{c}	n n$^{j\ c}$	n –^{c}
	BN	ɣ	ɣ^{ɦ}	k	x	m	m	n	n	n	ŋ

[a] [u_a] (if not given, the environment [i_a] may be missing from the data).
[b] [i_a] (if not given, the environment [u_a] may be missing from the data).
[c] [i_a] and [u_a].

Throughout Table 4.9 /m^{ɦ}/ is phonetically [m̤], /n^{ɦ}/ is phonetically [n̤], /l^{ɦ}/ is phonetically [l̤], and /ɣ^{ɦ}/ is phonetically [ɣ̈].

Table 4.10: Summary of inherited consonants and their reflexes (intervocalic cont.)

*		*l	*ɭ	*l^{ɦ}	*r	*ʃ	*h	*w	*j
	KS	l	l	l^{ɦ}	r	s	Ø^{a} h	w	j
	RL	l	l	l^{ɦ}	r	s	h	w	j
	MH	l	l	l^{ɦ}	r	s	h	w	j
	TH	l	l	l^{ɦ}	r	ʃ		w	j
V_V	SH	l l;ʎ^{b}	l -^{b}	l l;ʎ^{c}	r;ɽ	ʃ ʃ;ʃ$^{j\ c}$	Ø^{d}	w	j
	RP	l ʎ;l$^{j\ b}$	l ʎ^{b}	l ʎ;l$^{j\ b}$	r r$^{j\ b}$	ʃ ʃ$^{j\ b}$	Ø$_{b}^{d}$ –	w w$^{j\ b}$	j
	BH	l ʎ;l$^{j\ b}$	l ʎ^{b}	l l$^{j\ b}$	r r$^{j\ b}$	ʃ ʃ$^{j\ b}$	Ø$_{b}^{d}$ –	w w$^{j\ b}$	j
	BN	l	l	l	r	s^{f} h		w	–

[a] #CV_V#
[b] [i_a] and [u_a].
[c] [i_a] (if not given, the environment [u_a] may be missing from the data).
[d] This correspondence is found in most Tadbhavas (native, inherited vocabulary), while /h/ is often maintained in Tatsamas (Sanskrit loan words).
[e] This correspondence is irregular. Even within Tadbhavas there seems to be no categorical pattern of either deleting or maintaining medial *h in BN's linguistic history.
[f] Correspondence for intervocalic *ʃ in BN is irregular. Some etyma have /s/ and others /h/ without phonological conditioning. This irregularityis the result of BN's mixed Asamiya-Kamta linguistic ancestry, see further §4.3.13 and §7.5.4.2.

Table 4.11: Summary of inherited consonants and their reflexes (final position)

*		*b	*b^{ɦ}	*p	*p^{h}	*d̪	*d̪^{ɦ}	*t̪	*t̪^{h}
	KS	b	b	p	ɸ	d̪	d̪	t̪	t̪^{h}
	RL	b p [a]	β	p	ɸ	d̪;t̪	d̪	t̪	t̪
	MH	b p [a]	b	p	ɸ	d̪;t̪	d̪	t̪	t̪^{h}
_#	TH	b p [a]	β	p	ɸ	d̪	d̪;t̪	t̪	t̪
	SH	p	ɸ	p	ɸ	t̪	t̪	t̪	t̪
	RP	b p [a]	β	p	p	d̪	d̪	t̪	t̪
	BH	b p [a]	ɸ	p	ɸ;p	t̪	t̪	t̪	t̪
	BN	b	ɸ	p	ɸ;p	t	t^{h}	t	t

[a] Generally in Persian words, see §4.3.2.

Table 4.12: Summary of inherited consonants and their reflexes (final position cont.)

*		*ɖ	*ɖ^{ɦ}	*ʈ	*ʈ^{h}	*ʤ	*ʤ^{ɦ}	ʧ	ʧ^{h}
	KS	r		ʈ	ʈ^{h};ʈ	ʤ		ʧ	ʧ
	RL	ɽ;r		ʈ	ʈ^{h}	ʤ		ʧ	ʧ^{h}
	MH	ɽ;r		ʈ	ʈ^{h};ʈ	ʤ		ʧ	ʧ^{h}
_#	TH	r		ʈ	ʈ^{h};ʈ	ʧ		ʧ	ʧ;ʧ^{h}
	SH	r		ʈ	ʈ	s		s	s
	RP	ɽ;r		ʈ	ʈ	ʣ		s	ʦ;s
	BH	r		ʈ	ʈ	s		s	ʦ^{h};s
	BN	r		ʈ	t;t^{h}	s		s	s

Table 4.13: Summary of inherited consonants and their reflexes (final position cont.)

*		*g	*g^{ɦ}	*k	*k^{h}	*m	*m^{ɦ}	*n	*n^{ɦ}	*ɳ	*ŋ
	KS	g	g	k	k	m		n		n	ŋ
	RL	g	g	k	k^{h}	m		n		n	ŋ
	MH	g;k	g	k	k^{h}	m		n		n	ŋ
	TH	g	g	k	k^{h}	m		n		n	ŋ
_#	SH	k	k	k	k	m		n		n	ŋ
	RP	g	g	k	k	m; ˜		n		n	ŋ
	BH	k	k	k	k	m		n		n	ŋ
	BN	g;k	x	k	x	m		n		n	ŋ

Table 4.14: Summary of inherited consonants and their reflexes (final position cont.)

*		*l	*l̥	*lʱ	*r	*ʃ	*h	*w	*j
	KS	l	l		r	s	Ø		
	RL	l	l		r	s			
	MH	l	l		r	s	Ø		
_#	TH	l	l		r	ʃ	Ø		
	SH	l	l		r	ʃ	Ø		
	RP	l	l		r	ʃ	Ø		
	BH	l	l		r	ʃ	Ø		
	BN	l	l		r	s	Ø		

4.3.1 Devoicing of word-initial stops

Bongaigaon stands apart from the other KRDS lects as having undergone the most radical changes in its consonant system. One of these changes is the devoicing of word-initial stops. Voiced aspirated obstruents (both stops and affricates) have undergone complete devoicing of the obstruent element in word-initial position, for example *bʱul > /pʱul/ ‘error, mistake’. This devoicing causes near homophony with words beginning in a voiceless aspirated consonant, for example /pʰul/ ‘flower’ as against /pʱul/ ‘error’. In general the homophony is not complete, and the phonemic distinction is maintained (between, for example, [pʱ] and [pʰ]) through breathy-voiced aspiration (consistently) and low tone (more variably). Figure 4.1 shows the pitch contours for a nearly homophonous minimal pair [pʰal] ‘ploughshare’ and [pʱa`l] ‘good’.

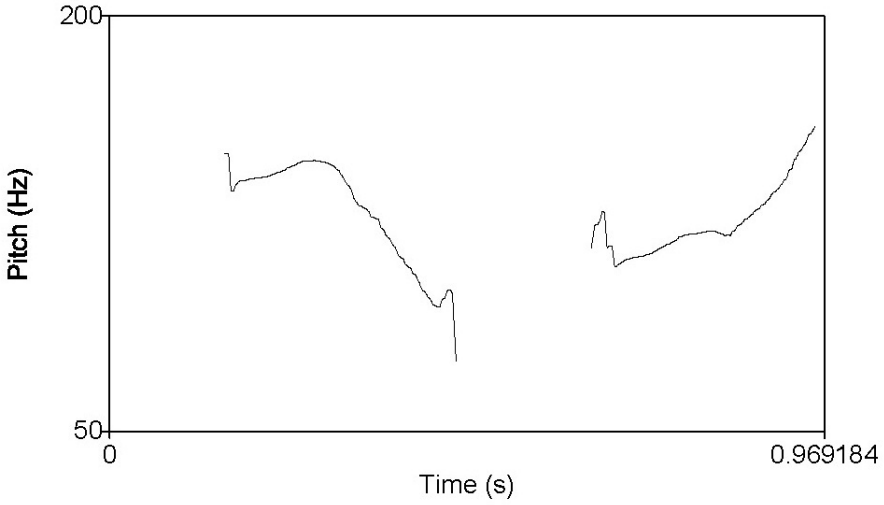


Figure 4.1: Pitch contours for [pʰal] ‘ploughshare’ and [pʱa`l] ‘good’

The stop element in both lexemes is voiceless [p], but the lexeme which has undergone devoicing (rightmost in the figure) has *low* (or perhaps low-rising) *tone*. The presence of low tone as a reflex of initial-position inherited voiced aspirates (e.g. *b^{ɦ}) recurs in the data, though it is somewhat variable.

The reflexes of protovoiced aspirates are, however, consistently distinguished in BN by *breathy-voiced* aspiration after the voiceless obstruent element. The quality of aspiration is thus the key feature which maintains the inherited phonemic distinction between voiced and voiceless aspirates: aspiration derived from inherited voiceless obstruents is 'clear' (e.g. *p^{h} > [p^{h}]), while aspiration derived from inherited voiced obstruents has breathy-voicing (e.g. *b^{ɦ} > [p^{ɦ}]). The presence of low tone is best explained as a phonetic accompaniment to breathy-voiced aspiration, and currently of no greater phonological significance.

This contrast between voiceless obstruents on the basis of differing qualities of aspiration is interesting, and perhaps unique in Indo-Aryan. Further observation of the BN consonant system should be undertaken to see whether this phonological arrangement is maintained, or whether tonal quality takes on phonemic significance. The breathy-voiced aspiration is transcribed in the data with /^{ɦ}/, as against clear aspiration /^{h}/.

It is not our concern here to consider other ways in which these consonant segments may be phonemicised (for example /p^{ɦ}/ versus /pV̤/), but such concerns should be addressed by further synchronic phonological studies. In particular, attention should be given to the articulation in BN of the inherited voiced aspirated affricate *ʤ^{ɦ} > [s̤], a breathy voiced fricative without aspiration.[8] This phoneme appears to be cross-linguistically significant, given Ladefoged and Maddieson's statement that 'There are no languages listed with breathy voiced fricatives' (1996:178). The concern of the present (diachronic) study is with the nature of the innovation—which is a phonetic devoicing change.

Other examples of the near homophony caused by this devoicing change—besides /p**^{h}**al/ 'ploughshare' and /p**^{ɦ}**al/ 'good'—include:

- /k^{h}ora/ 'lame, cripple' versus /k^{ɦ}ora/ 'horse'
- /p^{h}ɛra/ 'thigh' versus /p^{ɦ}ɛra/ 'sheep'
- /t^{h}akia/ 'having stayed versus /t^{ɦ}akia/ 'having covered'

This devoicing of initial voiced aspirates also occurs in Bhatibari, which out of the eight test sites is geographically closest to BN. Unlike for BN, however, the devoicing is variable in the BH data and occurs in a minority of possible occasions. Nonetheless, this devoicing is somewhat persistent in the BH data, being found on more than one occasion for all five of Bhatibari's voiced aspirates. It is most persistent in BH for initial *ʤ^{ɦ}. Examples from the BH data are:

- [b^{ɦ}ul ; p^{ɦ}ul] 'error'
- [d̤^{ɦ}uʌæ ; t̤^{ɦ}uʌæ] 'dust'
- [ɖ^{ɦ}ol ; ʈ^{ɦ}ol] 'drum'
- [ʣ^{ɦ}ori ; ts^{ɦ}ori] 'rain'
- [g^{ɦ}ɔnʈa ; k^{ɦ}ɔnʈa] 'bell'

---

[8] The relevant items in the BN wordlist data are: /s^{ɦ}ɔkɔra/ 'spear used for fishing' (no protoform reconstructed as part of this study), /s^{ɦ}aluk/ < *ʤ^{ɦ}aluk 'chilli', /s^{ɦ}ula/ < *ʤ^{ɦ}ula 'to hang', /s^{ɦ}ora/ < *ʤ^{ɦ}ora 'stream, small river'.

In Bongaigaon the devoicing goes further, to also affect *non-aspirated* stops in initial position. This change is variable in BN, but has high frequency. Unlike for the devoicing of aspirated stops in Bongaigaon (which is socially uniform), the devoicing of non-aspirated stops is probably conditioned by social variables. The uneducated speaker—a rickshaw puller—who recorded the wordlist gave devoiced variants on a majority of occasions, while the educated speaker gave voiced counterparts. Some examples:

- [**g**ai-ɣoru ; **k**ai-ɣoru] < *gai-goru ‘cow’
- [**b**an-pani ; **p**an-paɲi] < *ban-pani ‘flood’

Devoicing of *non-aspirated* stops is also found in the Bhatibari data, but highly infrequently. The geographical range coupled with the relative progress of the change in BN and BH suggest a change in progress (see Table 4.15), with propagation occuring from Bongaigaon towards Bhatibari. Further monitoring of the situation is required.

Table 4.15: Devoicing of initial stops in Bhatibari and Bongaigaon

	Bhatibari	Bongaigaon
Voiced Aspirated	Variable, minority, persistent	Regular
Voiced Unaspirated	Variable, non-persistent	Variable, majority, persistent

Looking around the NIA lects, a similar change can be found in some lects discontiguous with KRDS—most notably in the north-west of the sub-continent (e.g. Kashmiri [Koul 2003]), and to the south-east of KRDS in east Bengali dialects (e.g. Dhaka dialect [Pal 1966]). In all these lects, voiced aspirates have been lost from the phonemic inventory through devoicing, sometimes in coordination with deaspiration and the development of tone (Masica 1991:118–121, 204–205).

The geographical distribution of these structurally similar innovations in the extreme north-west and north-east corners of the NIA area prompts Cardona and Jain (2003) to propose contact with Tibeto-Burman languages as a causal condition for the development of this variation. This explanation would suggest that the devoicing change is not diagnostic of a propagation event because of its ecological non-distinctiveness (cf. §3.4.1.2).

However, the situation among the KRDS lects is different in important respects to the changes described above for other Indo-Aryan lects. In Bongaigaon and Bhatibari it is only in *initial position* that the voiced aspirates are devoiced—therefore there is no general merger between voiced and voiceless aspirates. Furthermore, as yet there is no reduction in the phoneme inventory even in initial position due to the maintenance of breathy-voiced aspiration despite the obstruent element losing its voicing. Lastly, as noted above, the devoicing of initial voiced aspirates in BN (and BH to a lesser extent) has been extended to the *unaspirated* voiced stops. For example, in BN we find regular *$b^{ɦ}$ > /$p^{ɦ}$/ word-initially, and variable but frequent *b > /p/ word-initially. Thus, while there is partial similarity with other innovations in Indo-Aryan, the conditioning is distinct enough from those cases to obviate any obvious link. On the grounds of this distinct linguistic conditioning, and in particular the development of a phonemic contrast based primarily on *aspiration quality*—which is unique, to my knowledge, in NIA—the initial devoicing of aspirated obstruents in BN (and BH variably) is diagnostic of a propagation event.

The innovations are summarised as follows:

[PI 1.] Devoicing of the obstruent element (not the aspiration) of initial voiced aspirates {regular in BN, variable in BH}. Diagnostic.

[PI 2.] Devoicing of initial obstruents {variable in BN}. Non diagnostic.

The sociohistorical conditioning of [PI 1.], involving BN (and BH peripherally) is investigated in Chapter 7.

4.3.2 Devoicing of word-final stops

The correspondences for word-final stops are brought together in Table 4.16 (reproduced from Tables 4.11 through 4.13) for ease of drawing comparisons across phonemes.

Table 4.16: Summary of correspondences for word-final voiced stops

*Phoneme	*b	*bʱ	*d̪	*d̪ʱ	*ʤ	*g	*gʱ
KS	b	b	d̪	d̪	ʤ	g	g
RL	b;p	β	d̪;t̪	d̪	ʤ	g	g
MH	b;p	b	d̪;t̪	d̪	ʤ	g;k	g
TH	b;p	β	d̪	d̪;t̪	ʧ	g	g
SH	p	ɸ	t̪	t̪	s	k	k
RP	b;p	β	d̪	d̪	ʤ	g	g
BH	b;p	ɸ	t̪	t̪	s	k	k
BN	b	ɸ	t	tʰ	s	g;k	x

In RL, MH, TH and RP there are occasional instances of devoicing of final *b. The same pattern is noted by Chatterji for SCB and SCA: 'Persian words in some cases show [p] for [b]' (1926:446). Indeed, the items in the collected wordlist that show *b > p in RL, MH, TH and RP are of Persian or Arabic origin: *ʤɔwab 'answer' (Arabic), *ʧɔrɔb 'fat (grease)' (Persian), *kʰarab 'evil, wicked' (Arabic).[9]

The data in Pattanayak (1966) show occasional final devoicing of *b in Perso-Arabic words for Oriya also (e.g. /kʰɔrapɔ/ 'bad, evil', with a later suffixed -ɔ). Note that *mɔt̪lɔb 'meaning' (Arabic origin) and *gɔrib 'poor' (also Arabic origin) occur without final devoicing in RL or MH, and in the case of *gɔrib in BH also. Final voicing in *mɔt̪lɔb suggests a recent borrowing of this item into RL and KS from Hindi; the etymon is not found elsewhere in the KRDS data, and these two lects exhibit the greatest Hindi influence of the eight surveyed here (cf. *ɔ unrounding, and changes in nominal morphology). In the case of *gɔrib, the etymon is found elsewhere in KRDS and cannot be a recent borrowing in the same class as *mɔt̪lɔb. A recent re-borrowing of *gɔrib via Hindi /gərib/ is sociohistorically plausible in the case of RL and KS, but not in the case of BH. No further conclusion can be reached at present on the irregularity of final *b > /b/ in BH /gorib/ 'poor'.

[9] The RP data has one exception, retaining the voicing in *ʧɔrɔb 'grease'.

Is this final devoicing of *b in Perso-Arabic words diagnostic of a propagation event and an erstwhile PN? Possibly, but a proper reconstruction of the scope of propagation and its chronology requires examination of Perso-Arabic loanwords in Oriya, Bangla, Asamiya, and further afield in NIA, and is thus beyond the scope of this study.

The evidence for devoicing of final *d̪ and *g in MH, and *d̪ʱ in TH, is scanty (one etymon for each) and non-conclusive.

The most persistent final devoicing is across SH, BH and BN. The relevant collected data are presented in Table 4.17 (with glosses in Note b).

Table 4.17: Analysis of final devoicing in SH, BH and BN [a]

	SH		BH		BN	
	-vc	+vc	-vc	+vc	-vc	+vc
b	zowap,[b] zip, bʱap, gorip	–	ʤɔp, ʤɔwap	gorib, pub	–	pub, gorib
bʱ	loɸ	–	loɸ	–	loɸ	–
d̪	nɔnɔ̪, oɪ̪, mɔ̪, ʃoa̪t, bipɔ̪		nɔnɔ̪, noɪ̪, mɔ̪, kʰɛ̪	–	howat, rot, hoat, bipɔt, amot	hrɔd; rɔd
d̪ʱ	ɖu̪	–	ɖu̪		dutʰ	
ʤ	ʃuɽus zahaɪ̪, ʃoβus, ʃans, moxɔs	–	ʤahas, ʃas, ʤɔmos	–	sahas, las, hɔhɔs, tɛs	–
g	rak	–	bʱok, nɔk, rak, ʃouk	–	rak; rag tʰɔk; tʰɔg	
gʱ	mɛk, bak, mak	–	mɛk, bak, mak	–	mɛx, bax, max	–
Total [c]	20	0	18	2	16	4
% -vc.	100%		90%		80%	

[a] A datum has not been considered an instance of final devoicing if the voice quality is possibly the result of assimilation with the initial consonant of a subsequent word, e.g. in Bhatibari: /kʰub bʱal/ ‘very good’, but /kʰup ʃuroɪ̪/ ‘very beautiful’.

[b] /zowap; ʤɔwap/ ‘answer’, /zip/ ‘life’, /bʱap/ ‘love’, /gorip, gorib/ ‘poor’, /pub/ ‘east’, /ʤɔp/ ‘barley’, /loɸ/ ‘temptation, /nɔnɔ̪/ ‘husband’s elder sister’, /oɪ̪, noɪ̪, rot/ ‘sunshine’, /mɔ̪/ ‘alcohol’, /ʃoa̪t/ ‘taste’, /bipɔ̪; bipɔt/ ‘danger’, /kʰɛ̪/ ‘sorrow’, /amot/ ‘enjoy’, /hrɔd; rɔd/ ‘lake’, /ɖu̪, dutʰ/ ‘milk’, /ʃuɽus/ ‘sun’, /zahaɪ̪, ʤahas, sahas/ ‘ship’, /ʃoβus/ ‘green’, /ʃans, ʃas/ ‘evening’, /moxɔs/ ‘brain’, /ʤɔmos/ ‘twin’, /las/ ‘shame’, /hɔhɔs/ ‘ease’, /tɛs/ ‘loud’, /rak; rag/ ‘anger’, /bʱok/ ‘pain’, /nɔk/ ‘vein’, /ʃouk/ ‘all’, /tʰɔg; tʰɔk/ ‘lie (untruth)’, /mɛk; mɛx/ ‘cloud’, /bak; bax/ ‘tiger’, /mak; max/ ‘month of the Hindu calendar’.

[c] The total is based on one point for regular occurrences and half of a point for variable occurrences (only relevant to the *g in BN).

Final devoicing of *b > p is not found in the data for BN, but the Perso-Arabic words generally devoiced in Eastern Indo-Aryan (see above) are also notably absent from the BN data. A larger or targeted wordlist for BN may yet provide some instances of final *b > p devoicing in lexemes of Perso-Arabic origin.

For SH, BH, and BN, final devoicing is regular for *d̤, *d̤ʱ, *ʤ, *bʱ, *gʱ.[10] The aspirated affricate *ʤʱ does not occur in word-final position. Devoicing of final *g is also found in these three lects, but less conclusively given the variation in BN and only one reflex of final *g in SH.

Final devoicing in SH, BH and BN does not closely parallel the initial devoicing changes (§4.3.1): final devoicing is largely uniform across SH, BH and BN; conversely, initial devoicing is advanced in BN, in progress in BH and not present (to any noticeable degree) in SH. Thus final devoicing and initial devoicing are here analysed as distinct changes.

Final devoicing of stops is not common in Indo-Aryan, but if we consider broader linguistic typology, the change is quite common. This change will be considered as supportive, but not in itself diagnostic, of a propagation event.

[PI 3.] *d̤, *d̤ʱ, *ʤ, *bʱ, *gʱ > [-voice] / _# {SH, BH, BN} (after rhoticisation).
Supportive, not diagnostic.

Two further changes are not diagnostic of propagation events:

[PI 4.] *b > [-voice] / _# {SH}. Diagnostic value unclear.

This change overlaps in linguistic conditioning with the devoicing of final *b in Perso-Arabic words (see above), and its diagnostic value is unclear.

[PI 5.] *g > [-voice] / _# {SH, BH, variably in BN}. Diagnostic value unclear.

This change may be diagnostic of a propagation event (perhaps even to be incorporated within [PI 3.]), but the data available to this study include only one reflex of final *g in SH, and variable of reflexes in BN, and hence are not conclusive.

The absence of *ɖ̤ from this set of final devoiced consonants suggests that *ɖ̤ had already undergone rhoticisation > [ɽ̤] > /r/ (cf. §4.3.8) before these final devoicing changes were propagated. That is, rhoticisation preceded final devoicing.

4.3.3 Palatalisation and labialisation of consonants

Palatalisation and labialisation of consonants are linguistically related changes which occur in central KRDS lects. This use of the term ‘palatalisation’ should not be confused with the MIA process of the same name, by which *ty, thy > cc, cch* (Bubenik 2003:218). In Rangpur and Bhatibari, a consonant occurring between a high vowel and the low vowel /a/ is palatalised: *C > C^{j} / V[+high] _ a {RP, BH}.

In Shalkumar the change is slightly different: palatalisation occurs after /i/ and labialisation after /u/, both still before /a/. That is, *C > C^{j} / i _ a {SH}, and *C > C^{w}/ u _ a {SH}. These changes are more variable in SH than in BH and RP, though nonetheless persistent in SH.

---

[10] With the exception of /hrɔd; rɔd/ ‘lake’ in BN—a *Tatsama* (recent Sanskrit loanword).

Table 4.18: Examples of labialisation and palatalisation across KRDS lects

	*ʤ > [ʤʲ] / i _ a	*p > [pʷ] (> [pʲ]) / u _ a
	'wet, damp' *bʱiʤa	'silver' *rupa
KS	bʱiʤa	-
RL	bʱiʤal	rupa
MH	bʱiʤa	rupa
TH	bʱiʤa	rupa; upa
RP	bʱizʲa	upʲæ
SH	bʱiʤʲæ	upʷa
BH	bʱiʤʲæ	rupʲæ
BN	pʱiza	rupa

The palatalised feature after the high-front vowel /i/ is linguistically natural, though apparently unique to this area within Indo-Aryan. Palatalisation after the *high-back* vowel /u/, as in RP and BH, is linguistically less expected than the *labialisation* attested for SH. Accordingly, the most linguistically plausible explanation involves reconstructing two changes: palatalisation of consonants /i_a/; and labialisation of consonants /u_a/. Subsequently the palatalised feature was generalised in both environments for RP and BH, supplanting the labialised feature. This account, which invokes linguistic seriation based on *plausible* diachronic dependency (cf. §3.4.3.1) implies three phonological changes:

[PI 6.] *C > C^{j} / i _ a {SH, RP, BH} (co-occurent with [PI 7.]). Diagnostic.

[PI 7.] *C > *C^{w} / u _ a {SH, RP, BH} (co-occurent with [PI 6.]). Diagnostic.

[PI 8.] *C^{w} > C^{j} {RP, BH} (after [PI 7.]). Diagnostic.

The palatalisation and labialisation changes described above are innovations which are uncommon among NIA lects. The conditioning environments are also relatively complex, and the range of propagation is contiguous and hence sociohistorically plausible as a PE. Accordingly, these changes constitute diagnostic evidence for reconstructing propagation events. A more detailed account of the ranges of these innovations is given in §7.4.1.1 where chronology of the changes is reconstructed by sociohistorical sequencing.

4.3.4 Deaspiration of medial consonants

Deaspiration in non-initial position is common in Bengali (Chatterji 1926:442), and in some of the KRDS lects. However, the discussion here of KRDS deaspiration will be with some limitations. The environments considered here exclude final position because the reconstruction is based on wordlists, without lengthy phonological study of most of the lects. The problem presents itself then of distinguishing in wordlist data between phonemic final *aspiration* and phonetic final *stop release*—be it aspirated or not. The presence or absence of final aspiration in KRDS lects is therefore left for further research. Whatever the result, it is unlikely to be diagnostic of a unified propagation event, because final

deaspiration is found in a number of NIA lects, most pertinently Bengali and Nepali. The contact through diglossia with both these languages in different KRDS areas increases the possibility of independent propagation, negating the likelihood of an integrated KRDS propagation event in this case.

However, the deaspiration of *medial* consonants is not so common in NIA—in particular it is not found in Nepali—making it more useful for diagnosing PEs (as well as being more reliably analysed from the wordlist data than final deaspiration). Medial deaspiration was found by Pattanayak (1966:62ff.) to be of historical importance for reconstructing '*AB'—a putative common stage in the Asamiya-Bangla linguistic histories. We will return to this broader picture after considering the prevalence of the change in our eight KRDS lects. The relevant correspondences from Tables 4.7 to 4.10 are reproduced below in Table 4.19 but without the footnotes to specify more detailed environmental conditioning. Medial *h is also included in Table 4.19, alongside aspirated consonants, for reasons given below.

Table 4.19: Summary of reconstructed medial aspirates and their intervocalic reflexes

*	*pʰ	*bʱ	*t̪ʰ	*d̪ʱ	*ʈʰ	*ɖʱ	*ʤʱ	*ʧʰ	*kʰ	*gʱ	*lʱ	*nʱ	*mʱ	*h
KS	ɸ	β	t̪ʰ	d̪ʱ	ʈʰ	rʱ	-	ʧʰ	kʰ;x	gʱ	lʱ	nʱ	mʱ	Ø h
RL	ɸ	βʱ	t̪ʰ	d̪ʱ	ʈʰ	ʈʱ;rʱ	-	ʧʰ	kʰ	gʱ	lʱ	nʱ	mʱ	h
MH	ɸ	βʱ	t̪ʰ	d̪ʱ	ʈʰ	ɖʱ;rʱ	–	ʧʰ	x	gʱ	lʱ	nʱ	mʱ	h
TH	–	bʱ βʱ	t̪ʰ	d̪ʱ	ʈʰ	rʱ	ʤʱ	ʧʰ	x	gʱ	lʱ	–	mʱ	h
SH	–	β	t̪ʰ	d̪	ʈʰ	ɽ	z	s	x –	ɣʱ	l l;ʎ	–	m	Ø
RP	– ɸʲ	– bʲ	t̪ -	d̪ d̪ʲ	ʈ ʈʲ	ɽ;r rʲ	ʤ	s sʲ	k kʲ	g –	l ʎ;lʲ	n –	m	Ø –
BH	-	β;b bʲ	t̪ -	d̪ d̪ʲ	ʈ ʈʲ	ɽ;r rʲ	ʤ	s sʲ	x kʰʲ	g;ɣ –	l lʲ	n –	m	Ø –
BN	–	βʱ	tʰ	dʱ	tʰ	r	z	s	x	ɣʱ	l	n	m	h Ø

The aspirated sonorants which are reconstructed for proto Kamta are: *lʱ, *nʱ and *mʱ (see §4.3.10 for the argument). In intervocalic position, aspiration is lost from these three protosonorants, from the medially-rhoticised *ɖʱ (cf. §4.3.8), and from the affricate *ɖʒʱ. This change is found in Shalkumar, Rangpur, Bhatibari and Bongaigaon; elsewhere in KRDS the aspiration is maintained. These five phonemes form a natural phonological class: voiced continuants.

The four lects with regular deaspiration of voiced continuants *also lose medial *h in regular or semi-regular fashion*. Most of the irregularity in correspondences is found in the Tatsama vocabulary (NIA borrowings from Sanskrit), while in Tadbhavas the correspondences show mostly regular loss of medial *h.[11] Kishanganj also loses medial *h, but it is phonologically restricted to the environment #CV_V#.[12] For example: (from KS) /gu˜i/ < *guhi 'crocodile', /lua/ < *lɔwha 'steel'; as against /pəhar/ < * pahaḍ 'mountain', /bɛhan/ < *bihaɳɔ 'dawn'. As discussed in §4.2, /h/ (and we may assume also *h) refers to the phonemic characteristic of breathy voicing accompanied by other phonetic features determined by the adjacent vowels.

The loss of aspiration from voiced affricates, nasals, laterals and rhotics in medial position is phonologically related to the deletion of medial *h. In all cases, *breathy voicing in continuant consonants is changed to modal voicing.*

[PI 9.] *C[+breathy voice, +continuant] > [+modal voice] / V_V {SH, RP, BH, BN, Oriya, Asamiya, SCB} ([tentatively] after C16th, after rhoticisation). Diagnostic.

This change is diagnostic of a propagation event based on ecological distinctiveness—as noted above, deaspiration is not common in Indo-Aryan outside of the eastern NIA lects.

As this change includes lects with a written tradition—Bangla, Asamiya and Oriya—the reconstruction of chronology may also be informed by historical textual evidence. In the historical literature of Asamiya we find that '-h- remained in [Early Asamiya] generally to be lost in modern Assamese' (Kakati 1962:168). Similarly, 'In modern Oṛiyā, at least in the standard colloquial and northern dialects, an intervocalic -h- is elided as in Bengali. But in early Oṛiyā, this -h- is seen to be preserved in many instances' (P.C. Majumdar 1970:xxxiii). Medial *lʱ, *mʱ and *nʱ are also deaspirated in these same lects. Oriya is unlike the others in one respect: *ɖʱ does not merge with *ɖ in medial position, but the two remain as distinct phonemes. Chatterji (1926:159) describes a general 'tendency to disaspiration' in Bengali, Asamiya and Oriya:

> In the Eastern (Māgadhī) group of speeches disaspiration of intervocal and final stops is more or less common at the present day, but … not more than 300 years old.

Confusingly, this statement is not supported by Pattanayak's (1966) data for SCB, SCA or SCO. Pattanayak's data for Bengali include cases of non-initial deaspiration of *pʰ, *t̪ʰ, *ʧʰ, *kʰ, but in all cases the deaspirated stops are word-*final*, not medial, contrary to Chatterji's general statement. Pattanayak's data include only one piece of evidence in support of medial deaspiration of *ʈʰ. There are three Bengali words in his list which inherit medial *ʈʰ: /paʈʰano/ < *pɔʈʰ- 'send'; /oʈʰa/ < *uʈʰ- 'raise'; and /**laʈi**/ < *laʈʰi 'stick'.[13] Only the last of the three has undergone deaspiration. We may conclude that the merger of medial *ʈʰ and *ʈ is not a regular feature of SCB, unless Pattanayak's data are at this point reflecting an archaic orthography rather than common speech.

---

[11] Though BN has some unexplained irregularity, see note e to Table 4.10.

[12] /gʌm/ < *gɔhɔm 'wheat' is an example not covered by the generalisation expressed here.

[13] A further Bangla form given in Pattanayak's study, /aŋʈi/ 'ring' may also be relevant, but he is unable to give a protoform.

While there is conflicting evidence for a general deaspiration change (along the lines of [PI 10.] below) in Bangla linguistic history, the loss of aspiration in voiced continuants (along the lines of [PI 9.]) is conclusively established for Bangla, Oriya, Asamiya and those of the KRDS lects listed under [PI 9.]. Chatterji (1926:442) writes regarding the phonological history of *ɖ̤ʱ (rhoticised as /ɽ̤ʱ/) in Bangla:

> It seems in the early 16th century, voiced aspirated forms like পড় «påṛh-» *read* ... বাঢ়ে «bāṛh» *increases* … still obtained, although it is likely that the aspiration had become feeble. The voiced aspirates seem to have preserved the aspiration (in the West Central dialect [i.e. SCB—MT]) longer than the unvoiced ones, in both final and intervocal positions.

If the loss of aspiration described here for *ɖ̤ʱ is representative of a loss of aspiration in voiced continuants generally, then Chatterji's diagnosis of a post-16th century chronology applies to [PI 9.]. This chronology will be applied 'tentatively' to this change, with further confirmation sought in Chapter 7 from sociohistorical sequencing.

A more general deaspiration process has occurred in Rangpur and Bhatibari, as shown by the rows of shading in Table 4.19. In these lects, no aspiration occurs in medial position.

[PI 10.] Loss of aspiration in all intervocalic consonants {RP, BH, ?SCB, ?Oriya}. Diagnostic.

In some cases, the merger with inherited voiceless consonants is not entirely concluded—obscured by variable fricativisation in the bilabial and velar series (cf. §4.3.5). A general deaspiration occurs nonetheless, and is diagnostic of a propagation event on the basis of ecological distinctiveness (as in the case of [PI 9.] above).

4.3.5 Postvocalic fricativisation

As alluded to just above, bilabial and velar stops in medial position are frequently fricativised in KRDS. This generally allophonic process resembles the process of lenition which took place in MIA: stops became fricatives, then went on to be lost altogether in several Apabhraṁśa dialects (Bubenik 2003:219).

The generalised correspondences relevant to postvocalic fricativisation in KRDS are given in Table 4.20 for the labial, velar and 'palatal' (i.e. laminal) series of stops and affricates (reproduced from Tables 4.7 to 4.13).

Table 4.20: Summary of correspondences relevant to postvocalic fricativisation

		Bilabial				Lamino-(post)alveolar				Velar			
		*b	*b^{ɦ}	*p	*p^{h}	*ʤ	*ʤ^{ɦ}	*ʧ	*ʧ^{h}	*g	*g^{ɦ}	*k	*k^{h}
	KS	w	β	p	ɸ	ʤ	-	ʧ	ʧ^{h}	g	g^{ɦ}	k	k^{h};x
	RL	w	β^{ɦ}	p	ɸ	z	-	ʧ	ʧ^{h};s	g	g^{ɦ}	k	k^{h}
	MH	β	β^{ɦ}	p	ɸ	z	-	ʧ	ʧ^{h}	g	g^{ɦ}	k	x
	TH	b β	b^{ɦ} β^{ɦ}	p	–	ʤ	ʤ^{ɦ}	ʧ	ʧ^{h}	g; ɣ	g^{ɦ}	k	x
V_V	SH	β	β	p;ɸ p^{w}	–	z z^{j}	z	s	s	ɣ	ɣ^{ɦ}	x k^{w} k^{j}	x –
	RP	b b^{j}	– b^{j}	p;ɸ p^{j}	– ɸ^{j}	ʤ ʤ^{j}	ʤ	ts;s ts;s^{j}	s s^{j}	g –	g –	k k^{j}	k k^{j}
	BH	β β^{j}	β;b b^{j}	p p^{j} p^{j};p^{w}	–	ʤ ʤ^{j}	ʤ	ts ts^{j}	s s^{j}	g;ɣ –	g;ɣ –	k k^{j}	x k^{hj}
	BN	b w	β^{ɦ}	p	–	z	z	s	s	ɣ	ɣ^{ɦ}	k	x
	KS	b	b	p	ɸ	ʤ		ʧ	ʧ	g	g	k	k
	RL	b p	β	p	ɸ	ʤ		ʧ	ʧ^{h}	g	g	k	k^{h}
	MH	b p	b	p	ɸ	ʤ		ʧ	ʧ^{h}	g;k	g	k	k^{h}
	TH	b p	β	p	ɸ	ʧ		ʧ	ʧ;ʧ^{h}	g	g	k	k^{h}
_#	SH	p	ɸ	p	ɸ	s		s	s	k	k	k	k
	RP	b p	β	p	p	ʤ		s	ts;s	g	g	k	k
	BH	b p	ɸ	p	ɸ;p	s		s	ts^{h};s	k	k	k	k
	BN	b	ɸ	p	ɸ;p	s		s	s	g;k	x	k	x

Most commonly fricativised in medial position are the KRDS protophonemes *p^{h}, *b, *b^{ɦ} and *k^{h}. These are fricativised across all of KRDS—with the exception of RP, and regularity of fricativisation is absent also in BH. Significantly, RP and BH are the lects in which intervocalic deaspiration and palatalisation are most regular (§4.3.3–4.3.5). It seems likely that these other processes have interfered with the efficacy of fricativisation in RP and BH.

Next most commonly fricativised in medial position are *ʧ^{h}, *ʤ, *g, *g^{ɦ} and *ʧ. In general, fricativisation is resisted for *p and *k, with the exception of Shalkumar—the

KRDS lect with most extensive medial fricativisation—which has fricativised reflexes of all these protophonemes.[14]

The incidence of fricativisation is somewhat different in final position as opposed to the intervocalic position discussed above. Word finally, fricativisation is standard for *p^h and common for *$b^ɦ$, but not for *k^h. Otherwise, word-final fricativisation is restricted to the laminals in SH, RP, BH, BN, and to the velars in BN.

Besides KRDS, intervocalic fricativisation of labials and velars is present in SCB (Chatterji 1926; Ferguson and Chowdhury 1960) and SCA, though this process is mainly mentioned in connection with aspirate consonants. For example, 'Intervocally, the aspirated stops [in SCA] are more lenis than the unaspirated ones' (Goswami and Tamuli 2003:406). Given the presence of this fricativisation process in Bangla, Asamiya and KRDS the question arises whether it traces its heritage back to an early eastern Magadhan stage. Chatterji seems to suggest so when he proposes fricativised allophones of /g, d, b/ as part of the phonology of Māgadhi Apabhraṁśa (1926:258). However, elsewhere he suggests that intervocalic spirant pronunciation of /p^h, $b^ɦ$/ is as recent as the early 20th century (ibid.:442–443). At this stage both the chronology, and the diagnostic value, of these intervocalic fricative allophones is far from clear. Further determination of this issue is left to later studies. Such studies should also consider whether there is a possible Tibeto-Burman substratum to this phenomenon, noting its presence in Asamiya, eastern Bangla and KRDS.

4.3.6 Merger of dental and postalveolar apical stops

The merger of the inherited apical series—dentals and postalveolars—is a distinctive feature of Asamiya (Baruah and Masica 2001; Goswami and Tamuli 2003), and the same merger is found in the Bongaigaon data collected for this study.[15] This is an unexpected result. In his study of the local NIA lects of the west of Assam, D.N. Das (1990) describes a fully maintained 'retroflex'/dental distinction for the lects of old Goalpara district, which includes Bongaigaon. He is not alone in this statement (cf. U. Goswami 1970, 1974). However, the data collected for this study contradict their statements. The two series—dental and apico-postalveolar—are merged into an alveolar series in the data collected in Bongaigaon as part of this study.[16] Very occasionally there is a postalveolar or dental articulation but in those cases the distinction is often not correctly reinstated. That is, there are cases of an inherited dental articulation being variably pronounced as a postalveolar, and vice-versa.

Assuming D.N. Das' (1990) and U. Goswami's (1970, 1974) descriptions to be accurate—a good assumption given their native ears—the merger of dental and apico-postalveolar series is a recent change. It will be worth ascertaining whether Das collected the data for his PhD thesis some time before 1990, and also whether those data represent a conservative phonology of older speakers. Alternatively, there may exist different sub-dialects within the Bongaigaon area, some of which are more conservative and retain the dental/'retroflex' distinction, others of which have merged the two series as described here.

---

14 *ʤ^ɦ occurs in medial position in only one etymon in the reconstruction: *maʤ^ɦ- 'between, middle'.

15 It is usually claimed to be unique in Indo Aryan to the Assam region, however Ed Boehm reports the merger also for one of the Tharu lects (pers. comm.).

16 For example, /pat/ < *paṯ- 'leaf', /pɛt/ < *pɛṭ 'stomach', /tatka/ < *ʈaʈka 'fresh', /taratari/ < *ṯara-ṯari 'quick, swift'.

Given the differences between previous descriptions and the description outlined here, the dental/'retroflex' merger may have occurred in BN as recently as the 20th century.

However, in SCA this merger is considerably older. Chatterji notes that in Asamiya written records from the 15th Century 'Assamese traits are occasionally noticeable: for example, the confusion between dentals and cerebrals [i.e. postalveolar apicals—MT]' (Chatterji 1926:108). On this evidence, the chronology of [PI 11.] for SCA may be dated at latest to the 15th Century, and plausibly earlier still. The propagation of the same change into BN, on the other hand, is a much more recent event.

[PI 11.] Apical series > alveolar articulation {BN and Asamiya} (during or before C15th in SCA, C20th in BN). Diagnostic of contact relations with SCA through diglossia.

This change has entered BN as a result of a diglossia with SCA.

4.3.7 Merger of dental and postalveolar nasals and laterals

From MIA and earlier NIA literature, as well as comparative data from modern NIA lects (such as Marathi and Oriya), we know that a similar distinction to that described above—between dental and postalveolar—pertained previously for nasals and laterals. None of the present-day KRDS lects maintain this inherited distinction between /n/ and /ɳ/, /l/ and /ɭ/. Pertinent facts from the historical textual record are summarised as follows (from Chatterji 1926:523 ff., 538ff.):

- In Late Middle Bengali texts (AD 1500–1800) there is always confusion between ণ (postalveolar nasal symbol) and ন (dental/alveolar nasal symbol);[17]
- There is some confusion between these two symbols in the *Sri Krishna Kirtana* (1300–1500) and the *Caryapadas* (earlier still). However, the incidence of ণ (postalveolar nasal symbol) is much higher, and confusion with ন (dental/ alveolar nasal symbol) considerably lesser, during the early Middle Bangla period than in late Middle Bangla;
- Chatterji concludes that 'It seems likely that Bengali possessed [the alveolar/postalveolar distinction in nasals] in the early Middle Bengali period' (*ibid.*), that is AD 1300–1500. 'From the beginning of the 15th century, probably, it ceased to exist as a cerebral' (ibid.);
- Modern Oriya /ɳ/ and /ɭ/ correspond to OIA medial /ɳ- n/ and /ɭ, l/ respectively. The postalveolar retraction of OIA single (i.e. not clustered) /n/ and /l/ is a MIA change reported by the Prakrit grammarians. However, during late MIA (Apabhramsa) the innovative postalveolar nasal and lateral returned to an alveolar articulation in initial position. Since late MIA therefore, an alveolar/postalveolar distinction has been maintained (if at all) in medial position only. Although 'Modern Oṛiyā is pretty definite [in maintaining this distinction], Middle Oṛiyā spelling, as in the 15th and 16th century inscriptions […], is not fixed in this matter' (ibid.);

[17] There is no distinct symbol for a postalveolar lateral in the Bangla-Asamiya script. This makes it harder in the case of laterals to date the loss of an alveolar/postalveolar distinction using textual evidence.

- This loss of apical distinctions among nasals and laterals is not unique to Bangla and Asamiya, but also found in the Bihari and Hindi lects. The maintenance of the dental/postalveolar distinction is attested to some degree by the early Maithili works including Vidāpati's Padāvalī (Jha 1985 [1958]:183). Vidāpati lived from the end of the 14th to the beginning of the 15th century AD (Yadav 2003), which suggests that the loss of this distinction in the Maithili area occurred sometime during or after the 15th century, as was suggested for Bengali above.

Given this evidence from historical texts, I tentatively conclude that the alveolarisation of nasals and laterals in Bangla and Maithili became regularised in these lects only after the 15th century. If this change occurred in KRDS lects at a similar or slightly later time to Maithili and Bangla, then the postalveolar nasal and lateral would have been present (at least variably) during the proto Kamta stage of development, AD 1250–1550 (cf. §7.3.1).

[PI 12.] *ɳ̣ , *ɭ̣ > /n, l/ {KRDS, SCB, SCA, Maithili, Hindi, etc.} (C15th or later). Possibly diagnostic based on sociohistorical plausibility (cf. §7.4.4), but broader NIA reconstruction is necessary in order to verify the diagnostic value of this change.

Whether or not the propagation of this structural change was interconnected between early Bangla, Asamiya, KRDS, Hindi, Bihari, etc., is a matter that may be addressed by a broader comparative study (though cf. §7.4.4).

In order that reconstruction of the proto Kamta lexicon be as realistic as possible, nasal and lateral postalveolars are included in the set of protophonemes in Table 4.30. They are also included in the KRDS comparative wordlist (Appendix A of Toulmin 2006) based on the following criterion: wherever a reconstructed nasal or lateral corresponds with Oriya /ɳ̣/ and /ɭ̣/ (which maintain the earlier NIA distinction), we can assume that the proto Kamta phoneme likewise had postalveolar articulation.

4.3.8 Rhoticisation

Rhoticisation of *ɖ̣ and *ɖ̣ʱ occurs in the environments: intervocalic, postvocalic word-final, and before or after non-apical consonants (cf. Wilde 2008:25; Masica 1991:97). Rhoticisation is innovative, but also widely distributed in NIA. Masica states: 'In much of NIA, MIA *-ḍ-* … became, at least allophonically, [*-ṛ-*]' (1991:194, here *ṛ* would seem to represent a postalveolar or retroflex tap [ɾ,ɽ]). He continues: 'In the Bihari and Eastern (and even some Western) Hindi dialects, Nepali, Assamese, and East Bengali dialects (and partly in Kashmiri also), this [*-ṛ*] has merged, no longer allophonically, with /*r*/'. The widespread range of this rhoticisation and alveolarisation in Eastern and Midlands NIA, as well as Nepali, complicates the reconstruction of PEs. The changes fail the diagnostic of ecological distinctiveness, and with no great complexity to commend themselves, the rhoticisation (and consequent alveolarisation) change is not diagnostic of an interconnected propagation event.

[PI 13.] *ɖ̣, *ɖ̣ʱ (> *[ɾ], *[ɾʱ]) > /r/, /rʱ/ / V_ {Bihari lects, several KRDS lects, SCA, Nepali, some Bangla lects, some Hindi lects}. Non diagnostic.

Some relative chronology for rhoticisation in KRDS can be established based on linguistic and textual seriation. However, given that we cannot reconstruct unified propagation events for rhoticisation, such sequencing is of little utility. Rhoticisation in

KRDS must have preceded changes in *final voicing* ([PI 3.]–[PI 5.]), for otherwise the devoicing change would bleed the rhoticisation, and the word final reflex of *ɖ would be /ʈ/ (which it is not). Secondly, rhoticisation must have preceded the merger of dental and postalveolar apical stops, because the alveolarisation change would bleed rhoticisation. The alveolarisation of apical stops is reconstructed above as [PI 11.] with chronology 'during or before C15 in Asamiya, C20 for BN'. Thirdly, rhoticisation—by turning *ɖʱ into a continuant—feeds the deaspiration of continuants by [PI 9.] in SH and BN and therefore rhoticisation must have preceded [PI 9.]. The deaspiration of continuants was reconstructed as prior to fricativisation of stops, and (tentatively) after C16th. However, unless we are able to diagnose the extent of propagation events involving rhoticisation, then this chronology cannot be put to use in reconstructing linguistic history.

4.3.9 The inherited 'palatal' (i.e. laminal) series of affricates

The reconstructed KRDS consonants include the typical Indo-Aryan series of obstruents (stops and affricates) at five places of articulation: bilabial, dental, apico-postalveolar, lamino-postalveolar ('palatal') and velar. These series are generally retained, but have undergone quite substantial restructuring in BN. Merger of the dental and postalveolar apical series has already been reconstructed in §4.3.6. In the laminal series of affricates also, changes take place which result in a restructured phonological system for BN.

Of the eight lects examined in this phonological reconstruction, Thakurgaon has the most strongly postalveolar articulation of affricates (e.g. [ʧ]). The series is slightly fronted towards the alveolar ridge in Kishanganj, Rangeli and Mahayespur, and is articulated on the ridge in Rangpur, Shalkumar and Bhatibari (e.g. [ʦ]). Especially in the central KRDS lects, the precise place of articulation of affricates is socially conditioned with more educated speakers favouring a more postalveolar pronunciation and less educated speakers favouring a more alveolar pronunciation. The situation is somewhat different in Bongaigaon where, similar to Asamiya, the inherited affricates have regularly become fricatives.

Fronting of affricates is not uncommon in Indo-Aryan, as Masica (1992:94) writes:

> There is a tendency in some languages and dialects to pronounce the /c/ as an alveolar (or 'dental') affricate [*ts*], e.g., in Nepali, Eastern and Northern dialects of Bengali (Dacca, Maimansing, Rajshahi), the Lamani and North-western Marwari dialects of Rajasthani, the Kagani dialect of 'Northern Lahnda', Kumauni, and many West Pahari dialects

Chatterji suggests in the case of North and East Bengali dialects that contact with Tibeto-Burman languages played a role in the phonetic adjustment of this series (Chatterji 1926:79). If this is the case, then this innovation fails the second diagnostic of 'ecological distinctiveness' (cf. §3.4.1.2), and thus it is not necessary that this change in articulation was propagated from north to east as Chatterji goes on to suggest. Rather, because Tibeto-Burman contact occurs independently in both north and east (as also in the Nepali speaking area), the possibility of independent development is quite high. Fronting of affricates to alveolar articulation is not, therefore, diagnostic of a propagation event.

4.3.10 The inherited aspirated sonorants

Aspirated sonorants (that is, rhotics, laterals, and nasals involving the coordination of modal and breathy voicing, cf. §4.2) are found sparingly and only in medial position in the four more western lects of this reconstruction: KS, RL, MH and TH. In the other four KRDS lects, as in SCB and SCA, aspiration is regularly lost in this environment by [PI 9.].

Aspirated and unaspirated pairs of stops are the norm throughout Indo-Aryan history, but the same opposition in the laterals and nasals is not as fundamental to the Indo-Aryan consonant system. For the phonemes *l^{ɦ}, *m^{ɦ}, *n^{ɦ} to be established as an inherited element of KRDS linguistic history, these phonemes must be found regularly in correspondences which are not explicable as innovations.[18]

Aspirated nasals and laterals formed part of the MIA phonology, as Bubenik (2003:206) writes:

> In MIA murmured nasals and the lateral liquid resulted through metathesis from OIA clusters *hm, hn, hṇ, hl: brāhmaṇa* 'brahman' > [Śaurasenī], [Māgadhī] *bamhaṇa*.

Though these phonemes are established for MIA, it remains to be determined that they were not lost in late MIA, with the modern aspirated nasals and lateral resulting by local innovations. This can be demonstrated by examining the aspirated correspondences in conjunction with OIA and MIA etymologies where known.

Table 4.21: Cognates with [l^{ɦ}] and their etymologies

English	'kite'	'now'	'yesterday, tomorrow'	'axe, hatchet'	'lame, cripple'
Etym.	*cilli* > *cillā* {Pk.}	vḗlā	*kālya*	*kuṭhāra*	*lulla*
Et. Id.	#327	#1210	#1511	#684	#990
pKmt	*ʧil^{ɦ}a	*ɛl^{ɦ}a	*kali	*kuɖ^{ɦ}alɔ	*(lula; *lul^{ɦ}a)
KS	ʧil^{ɦ}a	al^{ɦ}a	kal	kul^{ɦ}ari	–
RL	ʧil^{ɦ}a	al^{ɦ}a	kal^{ɦ}i	kur^{ɦ}al	–
MH	ʧil^{ɦ}a	al^{ɦ}a	kali; kəli	kul^{ɦ}ari	lul^{ɦ}a
TH	ʧil^{ɦ}a	ɛla; ɛlan^{ɦ}e; ɛl^{ɦ}aj	kail	kur^{ɦ}al	lula; nula
SH	tsiɲa	ɛla	kali	kuɽ^{w}al	–
RP	tsiɲæ	ɛla	kalk^{j}æ; kail	kur^{j}æl	nul^{j}æ
BH	tsil^{j}æ	æla	kali	kur^{j}æl	nul^{j}æ
BN	sila	ɛla	kali	kut^{h}ar [a]	–

[a] This form is a Tatsama that has come to BN as a loan from SCA.

[18] The aspirated stop *ɖ^{ɦ}—which undergoes lenition to a rhotic [ɽ^{ɦ}, ɾ^{ɦ}, r^{ɦ}] cf. §4.3.8—is excluded from the discussion in this section, as its inherited status, as part of the stop system, is not in doubt and the innovations that affect it have already been addressed above.

The evidence presented for inherited *l^{ɲ} in these five etyma varies in consistency. While #327 and #1210 attest /l^{ɲ}/ in corresponding positions for four lects (as indicated by the dark shading), the same pattern is not borne out by the other three etyma.

Aspiration in #1511 is restricted to Rangeli, though common with Maithili in this etymon (Jha 1985 [1958]:187). This formal match between Rangeli and Maithili, and not between Rangeli and the other KRDS lects, suggests the RL form to be a Maithili loan. This hypothesis is supported by the geographical position of Rangeli, and hence sociohistorically plausible. Of the eight KRDS lects examined here, RL is geographically closest to the Maithili speaking areas of Bihar and the Nepal Tarai. In his study of the Rajbanshi of Nepal, Wilde (2008:2) writes:

> Considering the centuries-old language contacts in south-eastern Nepal and the use of Maithili as a literary language and lingua franca (cf. also Jha 1958:28–29; van Driem 2001:1160–1161), it is hardly surprising that Maithili was to have an effect on the Rājbanshi varieties bordering the Maithili language area.

In #684, l^{ɲ} in KS and MH intrudes into an otherwise regular correspondence for *ɖ^{ɲ} across KRDS. The interchange of medial laterals and rhotics is found particularly in the history of Hindi and Bihari lects. The KS and MH forms are thus considered as loans from Hindi or Bihari.[19]

The distribution of /l^{ɲ}/ in #990 does not match any of the former correspondences. Lack of aspiration in TH conflicts with the correspondence for #327 and #1210. With aspiration present in only one lect it is ambiguous whether the aspiration in MH is inherited (with TH an irregularity or loan), or whether MH is a borrowing from Bihari (with TH representing the KRDS inheritance from MIA). The former explanation has some support, given that it is derived from OIA medial *ll* as too was #327 *cilli*. However, in the case of #990 the more authentically KRDS inheritance cannot be conclusively determined without systematically examining all the reflexes of OIA medial *ll* in TH and MH.

The firmest ground for an inherited aspirated lateral comes therefore from the correspondences in #327 and #1210. On the basis of the correspondences in these two cognates I tentatively reconstruct *l^{ɲ} as a proto phoneme of proto Kamta. Given this reconstruction, the correspondences are generally well explained by [PI 9.] (excluding loanwords) and the reflexes are the same as for aspirated nasals reconstructed below. Proto Kamta forms for the five etyma given in Table 4.21 are consequently reconstructed as: *ʧil^{ɲ}a ‘kite’; *ɛl^{ɲ}a ‘now’; *kali ‘yesterday, tomorrow’ (with /kal^{ɲ}i/ a Maithili loan); *kuɖ^{ɲ}al ‘axe, hatchet’; and *(lula; lul^{ɲ}a) ‘lame, crippled’. The data are insufficient at present to disambiguate whether one of these forms for ‘lame, crippled’ is a loan, or whether variability in this lexeme should be reconstructed for the proto Kamta stage.

---

[19] The Turner (66) entry for this etymon includes the following cognates: kuṭhāra 3244 kuṭhāra m. ‘axe’ ... Pa. kuṭhārī - f., Pk. kuḍhāra - m., kuhāḍa - m., ˚ḍī - f. (for ṭh - r ~ h - ḍ see piṭhara -), S. kuhāṛo m., L. P. kuhāṛā m., ˚ṛī f., P. kulhāṛā m., ˚ṛī f., WPah. bhal. kurhāṛi f., Ku. kulyāṛo, gng. kulyāṛ, B. kuṛāl, ˚li, kuṛul, or. kuṛāla, kurāṛha, ˚ṛhi, kurhāṛi, kuṛāri; Bi. kulhārī ‘large axe for squaring logs’; H. kulhāṛā m., ˚ṛī f. ‘axe’.

Turning to the evidence for inherited aspirated nasals, the cognates for /nʱ/ are less well distributed across KRDS, making it harder still to reliably reconstruct a proto Kamta phoneme. Aspirated nasals which correspond with a nasal-stop cluster in some lects—for example /banʱ-/ ‘to tie’ {MH} corresponding with /banḍ-/ ‘to tie’ {BH}—have not been included in Table 4.22 as in such cases the common KRDS inheritance is a *nasal-stop cluster*, and the aspirated nasal is an innovation (cf. §4.3.1.2)

Table 4.22: Cognates with [nʱ] and their etymologies

English	‘light’	‘milk’	‘elbow’	‘print, mark’
Etym.	*joṇhā* {Prakrit} < *jyṓtsnā* ‘moonlight’	*dṓhana*	*kaphōṇi*	*cihna*
Et. Id.	#208	#536	#389	#416
pKmt	*ʤonʱakɔ	*ḍunʱi	*kahuṇi	*ʧinʱɔ
KS	–	–	kənia; kʌnia	–
RL	ʤʌnʱak	ḍunʱi	konʱia	–
MH	ʤʌnʱak	ḍunʱi	kilkani [a]	ʧinʱʌ
TH	–	–	kɛheni	–
SH	–	–	kilkani [a]	sin
RP	–	–	kolkunʲæ; korkunæ [a]	–
BH	ʣonak	–	kilkæni [a]	–
BN	zonak; sonak	–	tilkani [a]	–

[a] This item is a partial cognate with the other KRDS forms. The cognate portion is -kani < *kahuni.

Analysis of the etymologies confirms Bubenik’s statement above that derived MIA *nh* from an earlier consonant cluster *hn*. The correspondence of nasals in #208, #536 and #416, while not well represented by cognate data across lects, nonetheless matches the MIA data well enough to justify reconstructing *nʱ as a proto Kamta inheritance. The aspiration is lost, as seen for *lʱ above, in the four more eastern lects (by [PI 9.]).

The aspiration in #389 is so restricted that it is more likely to have occurred through metathesis subsequent to the proto Kamta period. The etymon is reconstructed for proto Kamta as *kahuṇi > /kani/. Suffixation with *-a occurs in both KS and RL (cf. §4.4.11) and is followed by ‘shortening’ of *kahuṇi (> *kahnia) > /kʌhnia/ in KS by [PI 34.]. The corresponding vowel /o/ in RL, is not as predicted by [PI 32.], which would give /ɛ/ as in TH; the RL vowel in /konʱia/ is plausibly influenced by the Hindi form /kohni/ with the same meaning.

The aspirated bilabial nasal /mʱ/ is also quite rare in the wordlist data.

Table 4.23: Cognates with [mʱ] and their etymologies

English	'potter'	'blacksmith'	'to descend'
Etym.	*kumbhakāra*	*karm ̃āra*	*nam-*
Et. Id.	#828	#815	#1625
pKmt	*kumʱarɔ	*kamʱarɔ	*namʱ-
KS	kumʱar	–	namʱ-
RL	kumʱar	–	naβʱ-
MH	kumʱar	kumʱar [a]	numʱ- [a]
TH	kumʱar	kamʱar	–
SH	kumar	kamar	nam-
RP	kumær	kamar	nam-
BH	kumær	kamar	nam-
BN	kumar	kamar	nam-

[a] The vowel /u/ in these two items is not a printing error. The phonological difference with respect to the other KRDS forms is as yet unexplained.

Table 4.24: Cognates with [mʱ] (cont.)

English	'you. PL.ACC'	'them.PROX.PL .ACC'	'them.DIST. PL.ACC'	'you. PL.GEN'	'them.PROX. PL.GEN'	'them.DIST. PL.GEN'
Et. Id. [a]	#2303	#2305	#2306	#2320	#2323	#2324
pKmt	*t̪omʱa-[ʃa-]-kɔ	*[ɛmʱa-; imʱa-;ɛʃma-; iʃma-]-kɔ	*[omʱa-; umʱa-;oʃma-; uʃma-]-kɔ	*t̪omʱa-[ʃa-]-rɔ	*[ɛmʱa-; imʱa-; ɛʃma-; iʃma-]-rɔ	*[omʱa-; umʱa-;oʃma-; uʃma-]-rɔ
KS	t̪umsak	ismak	usmak	t̪umsar	ismar	usmar
RL	t̪əmʱak	jɛmʱa-k	amʱak	t̪əmʱar	jɛmʱa-r	amʱar
MH	t̪ʌmʱak	ɛmʱak	ʌmʱak	t̪ʌmʱar	ɛmʱar	ʌmʱar
TH	t̪umʱak	imʱa-k	umʱak	t̪umʱar	imʱa-r	umʱar
SH	t̪omak	imak	umak	t̪omar	imar	umar
RP	t̪omak	emak	omak	t̪omar	emar	omar
BH	t̪omak	imʲæ-k	umak	t̪omar	imʲæ-r	umar
BN	tomak; tumak	imak	tamak	tumar	imar	tamar

[a] Reconstruction of these forms is in §5.6

The best evidence for inheritance of *m^{ɦ} is the correspondence in #828, which also agrees with #815, and fits the same pattern as seen above for *n^{ɦ} and *l^{ɦ}—deaspiration in the four more eastern lects, maintenance elsewhere. The correspondences in #1625 also seem to be connected to *m^{ɦ}, but there is fricativisation and loss of nasal value in RL. The conditions of this lenition cannot be surmised from the data.

The aspirated nasals found in the pronouns would be straightforwardly harmonised with the other *m^{ɦ} correspondences, except that /m^{ɦ}/ in RL, MH and TH corresponds with /ms/ or /sm/ in KS. Several possible explanations need to be considered, all of them involving morphological changes, and these are more appropriately dealt with in Chapter 5 (cf. §5.6.1).

4.3.11 Changes to initial liquids

The initial liquids have a complex history in Eastern NIA, which traces back to the MIA period. The Eastern Prakrit underwent a complete merger of inherited /r/ and /l/ > /l/ (as attested by inscriptions and other sources [Masica 1991:186]). However, /r/ was re-established in contrast to /l/ due to the influence of Sanskrit as well as MIA varieties that had not undergone the merger. Chatterji finds the re-establishment of /r/ to have occurred before the 10th century, the point at which he holds Eastern Magadhan (Oriya, Bangla, Asamiya, KRDS) to have become differentiated from the other Magadhan lects (Chatterji 1926:537).

To add further complexity to the mix, MIA also saw interchange between /l/ and /n/ in initial position (Jha 1985 [1958]:187, who cites Pischel 1981 [1900]:§§260, 243). Chatterji writes that this interchange was inherited into Māgadhī Apabhraṁśa, and from there into the Magadhan lects (Chatterji 1926:545). However, the historical works on these lects show that this interchange is not regularly attested in modern Magadhan lects, and the primary reflex of OIA *l-* is /l/, not /n/ (Chatterji 1926:527ff., 543ff.; Jha 1985 [1958]:177ff., 187ff; Kakati 1962:229ff.; Tiwari 1960:67ff.). As in the case of the r/l merger, reintroduction of the distinction between initial l/n probably resulted from multilingualism during the Apabhraṁśa (late MIA) period.

With that introduction to the history of liquids in Eastern MIA and NIA, the generalised correspondences for reconstructed *l and *r in KRDS are as follows (reproduced from Tables 4.6, 4.10 and 4.14):

Table 4.25: Summary of reconstructed liquids and their reflexes

	#_			V_V			_#		
	*l,*ɭ	*lʱ	*r	*l,*ɭ	*lʱ	*r	*l,*ɭ	*lʱ	*r
KS	l		r n	l	lʱ	r	l		r
RL	l		r n	l	lʱ	r	l		r
MH	l		r n	l	lʱ	r	l		r
TH	l;n		r;Ø	l	lʱ	r	l		r
SH	n		r;Ø	l	l l;ʎ	r;ɽ	l		r
RP	n		Ø	l ʎ;lʲ	l ʎ;lʲ	r rʲ	l		r
BH	n		Ø;n	l ʎ;lʲ	l lʲ	r rʲ	l		r
BN	l		r	l	l	r	l		r

The reflexes for the aspirated lateral are not our concern here, as they have been addressed in §4.3.10. There are a few stray retentions of the Magadhan confusion between r/l (e.g. KRDS *ʃɔril ‘body’ derived from Sanskrit *śarir*; *[rɛʃun, lɛʃun] ‘garlic’ derived from Sanskrit *láśuna*). However, the majority pattern is the re-establishment of the older contrast in all positions.

The main changes to these protophonemes occur in initial position where *l > /n/ has been regularised in RP, BH and SH. While this change also occurred during MIA, the evidence from Eastern NIA lects discussed above shows that the l/n distinction was reintroduced during late MIA, as for the r/l merger. It is not uncommon in fact for the same change to re-occur independently in different stages of Indo-Aryan history, for example: medial spirantisation (cf. §4.3.5), rhoticisation of intervocalic /ɖ/, and so on. The strong regularity in the correspondence between /n/ in RP, BH and SH, and /l/ in the other varieties (excluding for the moment TH), is good evidence that the distinction had been thoroughly reintroduced by the proto Kamta stage, and that the nasalisation innovation is the result of a post-proto Kamta propagation event.

The *l > n change is found side by side with a change to the other inherited liquid, *r, in the very same varieties. Initial *r > Ø in Rangpur (regularly) and Shalkumar (variably), while in Bhatibari initial *r > n or Ø variably. Both these changes are socially stereotyped as ‘uneducated speech’ to a greater extent than *l > n. This awareness of a negative social stereotype seems to be the reason for irregularity, or variability, of *r > [r;Ø] in SH, and *r > [Ø;n] in BH. The change in initial *r is most regular in Rangpur (of the eight KRDS lects). The name of the town is accordingly pronounced by locals as [ompur].

Changes *l > n and *r > Ø are also found in Thakurgaon, but they are found predominantly in the speech of the local Hindus—Poliya and Rajbanshi—and not as frequently in the speech of local Muslims.

Given the irregular inheritance of MIA initial *l > n, there is some chance that re-occurrence could have happened independently in separate areas. In the south of West Bengal (Midnapore district), non-contiguous with the KRDS area, the same change is found amongst the Lodha people (D. Dasgupta 1978:156. For example [loha, noha] 'iron'). Therefore this change is not *diagnostic* of a propagation event because of the possibility of independent and parallel changes (given the MIA history). However, an examination of the range of the change in North Bengal shows it to be neatly contiguous in this area. Therefore, it is still possible that the change occurred through a propagation event, even if it cannot *diagnose* that propagation event. Thus [PI 14.] is 'supportive, but not diagnostic' of a propagation event. That is, if another change which *is* diagnostic of a propagation event (and hence a PN) shares the same range as [PI 14.], then [PI 14.] will be considered to have been propagated within the same network. This turns out to be the case, see §7.4.1.

[PI 14.] *l > /n/ / #_ {RP, SH, BH, and TH Hindus}. Supportive, not diagnostic.

The loss of the initial rhotic—in RP, SH, BH and amongst TH Hindus—and the nasalisation of the initial rhotic—variably in BH—are less common in eastern Magadhan than the l>n change. The ranges are contiguous, and these changes are diagnostic of PEs, despite not being particularly complex in their conditioning.

[PI 15.] *r > Ø / #_ {RP, variably in SH & BH, and among TH Hindus}. Diagnostic.

[PI 16.] *r > n / #_ {BH variably}. Diagnostic.

There is one further change left to be discussed in this section: initial *r > n in the lects KS, RL, MH in just one item of the collected data, *rand̪^{ɦ}- 'cooks' > /nan^{ɦ}-/. There are no other instances in the collected data of a protoword with the word-initial sequence *rVnC. It is linguistically plausible that this is the conditioning environment for the change of initial *r > n in {KS, RL, MH}. Given the close historical relationship between these three lects, it seems unlikely that this correspondence is a chance occurrence, but without further tokens to instantiate the correspondence it is not conclusive evidence for a propagation event.

4.3.12 Homorganic nasal + stop clusters

Medial clusters were highly common during OIA, but were generally assimilated to geminates, and then reduced to single stops during MIA (cf. Masica 1991:Ch.7). Sequences of homorganic nasal + stop, however, have been retained from OIA through to the reconstructed proto Kamta vocabulary. For example: *rand̪^{ɦ}- 'cooks' < {OIA} *randháyati* 'subjects'. Comparable forms are given in Table 4.26.

Table 4.26: KRDS forms with homorganic clusters of nasal + voiced aspirated stop

English	'cooks'	'wears'	'ties'	'crocodile'	'smells (perceives)'
Etym.	*randháyati*	*pinaddha*	*bándhana*	*kumbhīra*	**śr̥ṅkhati; śiṅghati*
Et. Id.	#1756	#1911	#1741	#255	#1698
pKmt	*randⁿ̪-	*pindⁿ̪-	*bandⁿ̪-	*kumbⁿirɔ	*ʃuŋgⁿ-
KS	nadⁿ̪-	pinⁿ-	bandⁿ̪-	–	suŋⁿ-
RL	nadⁿ̪-	pinⁿ-	banⁿ-	–	suŋⁿ-
MH	nanⁿ-	pinⁿ-	banⁿ-	–	suŋg-
TH	ranⁿ-	pinⁿ-	bandⁿ̪-	kumir	ʃuŋ-
SH	andⁿ̪-	pindⁿ̪-	bandⁿ̪-	–	ʃuŋ-
RP	andⁿ̪-; ɔndⁿ̪-	pendⁿ̪-	bandⁿ̪-	kumbⁿiɽ	ʃoŋg-
BH	randⁿ̪-	pendⁿ̪-	bandⁿ̪-	–	ʃuŋ-
BN	randⁿ-	pindⁿ-	bandⁿ-	–	huŋ-

Reduction of nasal stop clusters occurs in lects KS, RL, MH and TH, though the effects of this change are somewhat erratic. Reduction occurs for all four lects in item #1911, but the pattern is not repeated in other items. The repetition of irregular reduction in only these four lects does not seem to be by chance. However, the reflex is not sufficiently consistent to justify a propagation event. The change may *support* a PE diagnosed on other grounds.

[PI 17.] Homorganic cluster of N C [+asp, +vc] > N [+asp] {irregularly in KS, RL, MH, TH} Supportive, not diagnostic.

Within this same set of lects there is also a case of metathesis involving a homorganic cluster of *nd̪:

[PI 18.] *ɛnd̪urɔ > /nid̪ur/ 'rat' {KS, RL, MH, TH}. Supportive, not diagnostic.

There are no other instances in the data of the sequence #Vnd̪. The diagnostic value of [PI 18.] can only be supportive of a PE—to be reassessed based on a greater sampling of data.

4.3.13 The inherited sibilant

While OIA had three sibilants *s* (dental), *ś* (lamino-postalveolar) and *ṣ* (apico-postalveolar), these distinct phonemes were merged into a single sibilant in most MIA dialects (Bubenik 2003:216). In the Māgadhī Prakrit and Apabhraṁśa the single sibilant had *postalveolar pronunciation*, while elsewhere the pronunciation was alveolar. The postalveolar pronunciation, though now rare in Magadhan lects nonetheless constitutes a retention where it does occur, and thus is not diagnostic of propagation events.[20]

[20] Masica writes: 'In NIA the most widespread pattern consists of one voiceless sibilant, generally [s], plus /h/. In Standard Bengali, the dominant sibilant allophone is [ʃ] (becoming [s] before dental consonants). Although this is a Magadhan inheritance, it is not maintained in other modern Magadhan (Eastern NIA) languages (e.g. not in Assamese, Oriya, or 'Bihari')' (1991:98).

The correspondences for *ʃ across KRDS are as follows (reproduced from Tables 4.6, 4.10 and 4.14 without more detailed conditioning environments).

Table 4.27: Reflexes of the inherited sibilant in KRDS

	*ʃ		
	#_	V_V	V_V
KS	s	s	s
RL	s	s	s
MH	s	s	s
TH	ʃ	ʃ	ʃ
SH	ʃ	ʃ ʃ; ʃʲ	ʃ
RP	ʃ	ʃ ʃʲ	ʃ
BH	ʃ	ʃ ʃʲ	ʃ
BN	h s	s h	s

Postalveolar pronunciation in TH, SH, RP and BH is a retention, and not diagnostic of a PE. The palatalisation change (> [ʃʲ]) has been dealt with under §4.3.3. This leaves for consideration the anteriorisation (> /s/) in KS-RL-MH, and the anteriorisation and lenition in Bongaigaon (> /s/ > /h/).

Firstly, regarding the anteriorisation in KS-RL-MH. These lects are spoken in areas where the superposed lects are Standard Hindi and Nepali. Both these lects have an alveolar pronunciation for the sibilant. Contact through diglossia with these lects in Bihar and Nepal is a plausible explanation for the alveolar articulation in KS-RL-MH. The plausible role of diglossia in the change means that this shift in KS-RL-MH to alveolar pronunciation is not diagnostic of a propagation event. It does, however, support the diagnosis of contact relations (through diglossia) between KS-RL-MH and Hindi and/or Nepali.

[PI 19.] *ʃ > /s/ {KS, RL, MH (from Hindi, Nepali)}. Supportive, not diagnostic, of contact through diglossia with Hindi and/or Nepali.

In its treatment of the inherited sibilant phoneme, Bongaigaon once again differs from the other KRDS lects. Proto Kamta *ʃ becomes:

- (mostly) /h/ in word-initial position;[21]
- either /s/ or /h/ for different etyma in intervocalic position;[22]
- (mostly) /s/ in word-final position.

---

[21] Some of the exceptions can be explained as Sanskrit loans (e.g. /sɔkti/ ‘power’ < śakti and /sɔ̃sar/ ‘world, universe’ < saṁsāra); but some cannot (e.g. /hat; sat/ ‘seven’ < sapta).

[22] BN /s/ is, in other instances, the regular reflex of proto Kamta *ʧ or *ʧʰ, and in yet other instances a less regular reflex of *ʤ. See §4.3.9 for the reflexes of *ʧ and *ʧʰ, and §4.3.1–4.3.2 for the devoiced reflexes of *ʤ.

The lenition of *ʃ in BN bears partial similarity with Asamiya, in which the situation is as follows (Kakati 1962:63, fn.10):

- the MIA sibilant becomes /χ/ word-initially in both Tatsamas and Tadbhavas, and intervocalically but only in Tatsamas;
- the MIA sibilant becomes /h/ postvocalically in Tadbhavas.

The lenition in BN is plausibly connected with the Asamiya lenition, though the difference in conditioning environments means that the connection is not immediately transparent. Word-initially, the BN lenition (>h) follows the Asamiya model (>χ), but takes it one step further. Word-finally, BN does not follow the Asamiya model, but maintains the sibilant. The value of the BN changes for diagnosing PEs is not yet clear.

4.3.14 Approximants

The KRDS approximants /j/ and /w/ (also called semi-vowels) are uncommon segmental phonemes. In most of KRDS, as in SCB and SCA, their usual occurrence is as a glide to a diphthong. Occurrences of non-diphthongal [w] are often either the result of *b lenition, or of Perso-Arabic origin (e.g. ʤɔwab < *javāb* 'answer', ɖɛwal < *dīvār* 'wall'). However, there are further occurrences of both /w/ and /j/ which cannot easily be classed as vocalic. Relevant reconstructed protowords and their reflexes are given in Table 4.28.

Table 4.28: Cognates with reflexes of non-vocalic *j and *w

English	'heavens, sky'	'child'	'new'	'shadow'
Etym.	*div*	*śāvaka* 'fledgling'	*naviya-*	*chājā*
Et. Id.	#184	#792	#1340	#32
pKmt	*ɖɛwa	*ʧʰawa	*nɔja	*ʧʰa˜ja
KS	–	ʧʰua	nəja	ʧʰə˜ja
RL	–	–	nʌja	ʧʰa˜ha
MH	–	ʧʰua	nʌja	ʧʰi˜a
TH	–	ʧʰua	nɔja	ʧʰia
SH	–	sawa	nɔja	tsʰɛma
RP	ɖæwa	sawa	nɔja	sæ˜jæ; tsʰæ˜ja
BH	ɖɛwa	sawa	nɔja	tsʰɛŋa
BN	–	sawa	–	soja

The data in Table 4.28, along with others, justify the reconstruction of protophonemes *j and *w. They are generally retained in all KRDS lects, and so do not figure in the definition of any propagation event—though some phonologically irregular changes occur involving the approximants and adjacent vowels.

The third approximant found in KRDS (recall 4.2) is /h/. The reflexes for *h are as follows (reproduced from Tables 4.6, 4.10 and 4.14):

Table 4.29: Reflexes of *h in KRDS

	*h		
	#_	V_V	_#
KS	h	Ø h	Ø
RL	h	h	
MH	h	h	Ø
TH	h	h	Ø
SH	h	Ø	Ø
RP	h	Ø –	Ø
BH	h	Ø –	Ø
BN	h	h Ø	Ø

The regular reflex of *h in initial position is /h/ across all of KRDS. In medial position *h is deleted in some KRDS lects (see further §4.3.4).

4.3.15 Summary of reconstructed consonant system

Based on the phonological reconstruction contained in this chapter, Table 4.30 shows the reconstructed consonant system for proto Kamta after it split from the Māgadhī Apabhraṁśa. References to the sections that deal with relevant changes are given in italics.

Table 4.30: Consonant phonemes inherited into KRDS

	Bi-labial	Apical			Laminal Post-alveolar	Dorsal Velar	Laryngeal Glottal
		Dental	Alveolar	Post-alveolar			
Compare:	*4.3.5*	*4.3.6*		*4.3.6, 4.3.8, 4.3.7*	*4.3.5, 4.3.9*	*4.3.5*	
4.3.3	*p	*t̪		*ʈ	*ʧ	*k	
4.3.3, 4.3.4	*pʰ	*t̪ʰ		*ʈʰ	*ʧʰ	*kʰ	
4.3.1, 4.3.2, 4.3.3	*b	*d̪		*ɖ	*ʤ	*g	
4.3.1, 4.3.2, 4.3.3, 4.3.4	*bʱ	*d̪ʱ		*ɖʱ	*ʤʱ	*gʱ	
4.3.3	*m		*n	*ɳ		*ŋ	
4.3.3, 4.3.4, 4.3.10	*mʱ		*nʱ				
4.3.3, 4.3.13					*ʃ		
4.3.4.							
4.3.3, 4.3.11			*r				
4.3.3, 4.3.11			*l	ɭ			
4.3.3, 4.3.4, 4.3.10			*lʱ				
4.3.3, 4.3.14	*w				*j		*h

The justification for considering these phonemes to be a contemporaneous system (proto Kamta) stems from (i) the entire reconstruction of linguistic history which culminates in Chapter 7, and (ii) the post Kamta chronology reconstructed for the phonological changes mentioned above.

A summary of all the phonological changes diagnostic or supportive of propagation events (and thus pertinent to the reconstruction of speech community events in Chapter 7) is given at the end of this chapter.

4.4 Comparative reconstruction of KRDS vowels

The vowel systems of NIA lects classed as 'Magadhan' differ in two general ways from MIA and Midland NIA lects. Firstly, they differ by the loss of phonemic contrast between long and short vowels. This loss of contrast is found in Asamiya, Bangla, Bhojpuri (M. Verma 2003), Magahi (S. Verma 2003), Maithili (Yadav 2003), Oriya, as well as in KRDS. Secondly, the Magadhan NIA lects also show evidence of having inherited a backed articulation of the MIA short central vowel /ə/ > [ʌ, ɔ, ɒ]. This backed articulation is most strongly maintained in the Eastern Magadhan lects (Asamiya, Bangla, Oriya, KRDS), while the influence of Hindi /ə/ has partially eroded this articulation in Western Magadhan (Bihari) lects.

A summary of the reflexes of the proposed inherited vowel phonemes is given in Tables 4.31–4.34, with the reflexes of inherited vowel nasalisation in Table 4.35.

Table 4.31: Summary of inherited high vowels and their reflexes

	*i				*u		
	_(C) ɔ,o,ɛ	#C_C- [a]	$CVC_X\_C_YV$ where C_X or C_Y = l, n, r	else	_(Cx)ɔ,o,ɛ	#C_C- [a]	else
KS	i	i	i Ø	i	u	u	u
RL	i	i	i Ø	i	u	u	u
MH	i	i	i Ø	i	u	u	u
TH	i	i	i Ø	i	u	u	u
SH	i	i	i Ø	i	u	u	u
RP	e	e	Ø	i	o	o	u
BH	i	e	Ø	i	u	o	u
BN	i	i	i	i	u	u	u
see:	§4.4.3	§4.4.3	§4.4.6		§4.4.3	§4.4.3	

[a] Monosyllabic verb roots.

Table 4.32: Summary of inherited mid vowels and their reflexes

	*o				*ɛ			
	#_	$V_H(C)$_	_(C)a	else	#_	_CV_H	1st σ, else	else
KS	wo	o	ʌ	o	jɛ	i ɛ	ɛ	ɛ
RL	o	o	ʌ	o	ɛ	i ɛ	ɛ	ɛ
MH	o	o	ʌ	o	ɛe	i e ɛ	e ɛ	ɛ
TH	o	o	ɔ	o	ɛ	e	ɛ	e
SH	o	o	o	o	ɛ	e	ɛ	e
RP	o	u	o	o	æ	e	æ	e
BH	o	u	o	o	ɛ	e	ɛ	e
BN	–	–	o	o		e	ɛ e	e
see:		§4.4.2	§4.4.3	§4.4.1		§4.4.1		§4.4.4

Table 4.33: Summary of inherited *a and its reflexes

	*a				
	$\_C_Xu(C_Y)a$ [a]	$\_C_Xi(C_Y)a$ [a]	C_(C)CaC [b]	$V_H(C)$_	else
KS	ə	ə	ə	a	a
RL	ʌ	ɛ	ʌ[c]; a	a	a
MH	ʌ	ɛ	ʌ[c]; a	a	a
TH	a	o a ɛ	a	a	a
SH	a	a	a	a	a
RP	a	a	a	æ	a
BH	a	a	a	æ; ɛ	a
BN	a	a	a	a	a
see:	§4.4.7, §4.4.6		§4.4.7	§4.4.2	

[a] Where either C_X or C_Y is a sonorant consonant.

[b] This correspondence attests a change in the quality of *a in KS, RL and MH. In RL and MH the change is variable when the final consonant is a liquid; e.g. Rangeli /katʃʰar; kətʃʰar/ < *katʃʰar 'river bank'. In KS, where the change is most regular, it also occurs variably without a final C, i.e. in the environment C_Ca. For example /gəla/ < *gala 'cheek'; /tʃəna/ < *tʃana 'chick-pea'. See further §4.4.7.

[c] Varies phonetically with [ə].

Table 4.34: Summary of inherited *ɔ and its reflexes

	*ɔ				
	1st σ _CV$_H$	1st σ, else	_CV$_H$	_#	else
KS	ə [a]	ə	ə	Ø	ə
RL	ʌ [b]	ʌ [b]	ʌ [b]	Ø	ʌ [b]
MH	ʌ	ʌ	ʌ	Ø	ʌ
TH	o	ɔ	o	Ø	o
SH	o	ɔ	o	Ø	ɔ; o
RP	o	ɔ	o	Ø	o
BH	o	ɔ	o	Ø	o
BN	o	ɔ; o	o	Ø	ɔ
see:	§4.4.1		§4.4.5 §4.4.1		§4.4.4

[a] Some instances of [ʌ] but mostly the reflex is [ə].

[b] Varies phonetically with [ə].

Table 4.35: Summary of inherited vowel nasalisation and its reflexes

	*˜		
	mono-σ	1st σ of 2	2nd σ of 2
KS	˜	˜	Ø
RL	˜	˜	Ø
MH	Ø	Ø	Ø
TH	Ø	Ø	Ø
SH	Ø	Ø	Ø
RP	˜	Ø	Ø [a]
BH	Ø	Ø	Ø
BN	Ø	Ø	˜
see:	§4.4.10		

[a] Subject agreement endings on the verb are an exception to this generalised correspondence set, see §6.3.6

4.4.1 Regressive vowel raising

The main type of change in the history of KRDS vowels is vowel harmony, whereby specific qualities of a vowel are assimilated by neighbouring vowels. The vowel harmony found in KRDS, and in Eastern NIA more generally, involves *partial assimilation to the height value* of a nearby vowel.

In regressive vowel harmony, vowel qualities are assimilated in a regressive direction—from right to left, as it were. For example, the Old Oriya word for 'tamarind' is /t̪ɛnt̪ali/, but the modern Oriya equivalent is /t̪int̪ili/. The height of the final vowel /i/ has been assimilated from right to left (regressively), changing the features of earlier vowels in the word.

The opposite direction of assimilation is termed *progressive* vowel harmony, and involves a left to right direction of assimilation. For example, in RP and BH *ʃipa > /ʃipʲæ/ 'root'. In this case, the height of the first vowel conditions the raising of the latter vowel by one phonological 'notch' *a > /æ/. Here and in the text to follow, 'raising' and 'lowering' are used as shorthand for 'partial height assimilation' conditioned respectively by nearby high and low vowels.

Both progressive and regressive directions of vowel height assimilation are found in KRDS lects. Regressive vowel raising is described in this section, progressive raising in §4.4.2 and lowering (which is regressively conditioned) in §4.4.3.

In KRDS lects, regressive vowel assimilation affects *ɛ, *ɔ and *o, but not *a. The change is most widely distributed for *ɛ, more restricted for *ɔ, and sporadic for *o.

*ɛ > [e] / _ (C) V[+high] {MH, TH, SH, RP, BH, BN; SCB and SCA}[23]

*ɔ > [o] / _ (C) V[+high] {TH, SH, RP, BH, BN; SCB and SCA}[24]

I will now argue that the former of these two changes is an old allophonic change, and was inherited into proto Kamta from an earlier protostage, with the allophony subsequently lost in RL and KS.

In order to evaluate the chronology of regressive raising of *ɛ, consider the following comparative and textual evidence from various eastern Magadhan lects. In present-day Asamiya and Bangla, the reflex of inherited *ɛ is /e/ when a high vowel follows. While /e/ is always given the status of a phoneme in descriptions of Bangla and Asamiya, its phonemic status in these lects, distinct from /ɛ/, is quite marginal. Allophonic variation of /ɛ/ before a following high vowel can account for the vast majority of instances of /e/ in SCA words found in Kakati's (1962) index. The example given in Goswami and Tamuli (2003:77) as evidence for the e/ɛ phonemic distinction is: /bel/ 'bell' (an English loan) vs. /bɛl/ 'wood apple'. Indeed the status of /e/ in Asamiya as a distinct phoneme (rather than an allophone of /ɛ/) seems to be based largely on loans. A similar situation pertains in MH through borrowing from Hindi (see below). Likewise, the status of /e/ as a phoneme of SCB (as opposed to an allophone of /ɛ/) is bolstered by loans. However, in SCB and some KRDS lects, several instances of /e/ result from a morphologically-conditioned lowering of *i in verbal roots (cf. §4.4.3). This change produces minimal pairs in RP and BH such as /kene/ '(s/he) buys', /kɛne/ 'why'.

Irrespective of the present-day synchronic status of /e/, the instances of [e] in Bangla and Asamiya are derived for the most part from the vowel *ɛ inherited from the common Magadhan stage. Before English loanwords and before morphologically-conditioned lowering of *i, the sound [e] in earlier Bangla and Asamiya was an allophone of /ɛ/. However, there is no grapheme to mark this phonetic contrast in the Bangla-Asamiya script and therefore we would not expect to find evidence of allophonic [e] in historical written documents. What we do find is some evidence of regressive raising of *ɛ > *i* in Old Oriya:

---

[23] Sporadically raised as far as [i], e.g. *ʤɛʈʰi 'gecko' > /ziʈʰi/ {SH}.

[24] Sporadically raised as far as [u], e.g. *ʈokɔri 'basket' > /tukuli/ {BN}.

> The raising of the vowels -e- and -o- to -i- and -u- respectively is a notable feature in modern Oṛiyā. No doubt, the tendency had its origin in earlier times. In many instances of the earlier documents is to be found the old phonetic habit existing side by side with the new one. (P.C. Majumdar 1970:xxxiii)

Based on the argument given above, raising of *ɛ to [e] was originally allophonic in Bangla and Asamiya (and not marked orthographically). In support of this argument, raising of *ɛ to *i* is recorded variably in Old Oriya. It is reasonable to conclude that regressive vowel raising of *ɛ is old and possibly even an allophonic innovation inherited from an early common eastern Magadhan stage.

Given this hypothetical reconstruction, we must now consider how it can be that present-day RL and KS lack the regressive raising process. I propose here that proto Kamta inherited an allophonic regressive raising process of *ɛ > [e] before a high vowel; but because this process was allophonic, no inter-change of phonemes was entailed during or prior to the proto Kamta stage. It is possible that the allophony may have been lost in RL and KS, and have left little trace of the reversal of the older allophonic process. Based on this argument, the following change is reconstructed for all of eastern Magadhan lects.

[PI 20.] *ɛ > [e] / _ (C) V_H {eMg}. Diagnostic value unclear.

A proper assessment of the diagnostic value of this change will depend on historical reconstruction at the wider Magadhan level. This allophonic process was lost in RL and KS by [PI 21.].

[PI 21.] *[e] allophone of *ɛ > [ɛ] {RL, KS}. Not diagnostic.

The loss of allophony, similar to the loss of a variant discussed in §3.4.1.1, is of lower diagnostic value than the addition of an allophone, because of the possibility that the loss occurred independently and without interconnected propagation.

Regressive vowel raising of *ɛ is a phonologically general process, and distinct from the morphologically-conditioned raising (umlaut) which characterises SCB verbal morphology. Bangla umlaut has its origin in the duplication of /i/ before a preceding consonant. Chatterji terms this process ‘epenthesis’, and dates the change (based on written sources) as ‘well-established in all the dialects of Bengali by the beginning of the 15th century’ (1926:388). The anticipatory /i/ was later lost everywhere except in verbal morphology, where, the sequence *vowel + anticipatory /i/* was contracted to a single vowel possessing the qualities of the original vowel, but with raised height. Note, however, that Maniruzzaman (1977:35–36) cites Sukumar Sen (without immediate reference, perhaps Sen 1960 or 1971?) as having disputed this explanation of the origins of Bangla umlaut. The historical details of Bangla umlaut need not concern us overly here, the point is rather that some KRDS lects have anticipatory duplication, or ‘epenthesis’ along the lines of Middle Bangla, but the details are slightly different (see §4.4.6).

The other inherited phoneme affected consistently by regressive vowel raising is *ɔ.

[PI 22.] *ɔ > [o] / _ (C) V[+high] {TH, SH, RP, BH, BN, SCB and SCA}. Supportive, not diagnostic.

Here again, written sources are of only limited use in assigning a chronology to the innovation because the raising of *ɔ to /o/ before a high vowel is not reflected in current Bangla-Asamiya orthographic conventions. (There are distinct graphemes অ/ও for ɔ/o, but

the grapheme ɜ is not conventionally used for the higher vowel /o/ if that vowel results from /ɔ/ by regular regressive raising.)

Unlike for *ɛ, the raising of *ɔ is not attested in the present-day MH lect (with the one exception of /poxi/ 'bird' < *pɔk^{h}i). For example, the MH data include: /hʌrin/ 'deer' < *hɔrin; /nʌʈi/ 'throat' < *nɔʈi; /lʌḍi/ 'river' < *nɔḍi; etc. Furthermore, raising of *ɔ to /o/ is not reported for Oriya or Old Oriya. Crucial pieces of evidence which justified reconstructing inherited raising of *ɛ for all of KRDS (namely the presence of such raising in Old Oriya and MH) are absent for the regressive raising of *ɔ. Therefore, the change [PI 22.] is not reconstructed as part of the linguistic history of KRDS as a whole, but only for the subset of lects in which it is presently attested. Assuming that regressive raising was already present in these lects for *ɛ, the raising of *ɔ is not necessarily diagnostic of a unified propagation event because of the possibility of independent and parallel extension of regressive raising from the phoneme *ɛ to the phoneme *ɔ. Though the evidence may not be strong enough for [PI 22.] to be diagnostic of a propagation event in its own right, this change may constitute supporting evidence for the sociohistorical range of propagation {TH, SH, RP, BH, BN, SCB and SCA}—if this same range is established by some other more diagnostic change.

4.4.2 Progressive vowel raising

As stated earlier, both directions of vowel height assimilation are found in KRDS lects. Regressive vowel harmony has at least its roots in an inherited process, while progressive vowel harmony is a post-proto Kamta innovation. Under the progressive process, a vowel is raised one phonological notch by the presence of a preceding high vowel, as described below.

[PI 23.] *o > /u/ / V[+high]C_ {RP, BH}. Diagnostic.

[PI 24.] *a > /æ; ɛ/ / V[+high]C_ {RP, BH}. Diagnostic.

These individual outcomes of progressive raising have not been formulated as a single change because the dialectological range of propagation differs (see Chapter 7).

The protophonemes *o and *ɛ are rare in non-initial syllables within the reconstructed proto Kamta vocabulary.[25] The raising of *o>/u/ is reconstructed by comparison of the second person singular verbal endings (cf. §6.4.2.3).

Analysing the effects of progressive raising on *ɔ in RP and BH is not straightforward because *ɔ is regularly raised in non-initial position in RP and BH by prosodic vowel raising ([PI 28.], cf. §4.4.4). Thus raising of *ɔ > /o/ after a high vowel may equally be the result of progressive raising as of prosodic raising. All that can be said is that these two processes do not combine in any of the lects to give /u/ < *ɔ. (For example, /ʃoiʃo/ < *ʃɔiʃɔ 'maize', not ʃoiʃu).

Raising of *a is very frequent in RP and BH. For example, *ḍ^{ɦ}ula > /ḍ^{ɦ}ul^{j}æ/ 'dust'. It is linguisticially and historically related to the palatalisation and labialisation changes [PI 6.]-[PI 8.]. In the environments specified by those changes, the raising of *a > /æ, ɛ/ results in merger with *ɛ. This merger is regular and predictable in the speech of the less educated,

---

[25] It is possible that even in initial position *o may be progressively influenced by a high vowel in the last syllable of a preceding word, though this remains to be tested. Similarly, occurences of /e/ that result from regressive lowering [4.4.3]—e.g. /kene/ 'buys'—should be tested for susceptibility to raising when a high vowel ends the preceding word, e.g. /priti k[e~i]ne/ 'Priti buys'.

in village domains. The merger is generally less regular in Bhatibari-Tufanganj and Cooch Behar district in the speech of educated speakers, as well as in social domains where SCB has become dominant (such as urban life).

SCB has also undergone progressive raising of *a but the effects and conditioning of that raising differ to the process described here for the Cooch Behar and Rangpur areas of KRDS. The SCB process is summarised by P. Dasgupta (2003:358) as follows:

> The Bangla vowel harmony system also exhibits a counter-normal pattern, where a preceding trigger affects a right-hand eligible /ā/, with verb and non-verb subpatterns. The latter turns the target /ā/ … into a mid vowel copying the backness of the trigger: /bhije/ [[bhijā]] 'wet', /bhulo/ 'forgetful' … The subpattern for verbs turns /ā/ uniformly, if unexpectedly, into /o/.

By contrast, the KRDS change described here is phonologically general and without morphological restrictions. The KRDS change is distinct from the Bangla pattern, and also distinct from the inherited regressive pattern of vowel harmony in eastern Magadhan lects (cf.§4.4.1). As a result of this distinctiveness, changes [PI 23.] and [PI 24.] are diagnostic of propagation events.

4.4.3 Regressive vowel lowering

In addition to the raising of vowels before high vowels, some KRDS lects have also undergone regressive lowering of particular vowels under specific conditions. In western KRDS there is a general phonological lowering, and in the south-east of KRDS a different lowering process which is morphologically-conditioned in some lects, and phonological in others. The changes are not reconstructed as integrated between the west and the south-east because they operate on different vowels, under markedly different conditions.

Taking the western process first, the following etyma illustrate what is a mostly regular process of lowering in lects RL, MH and TH, and is also present (though obscured by Hindi loans) in KS. A few exceptions may also be found to this rule in RL and MH, but these are in a minority and do not take away from the general regularity of the change; they are likely to be the result of Hindi influence.

Table 4.36: Example correspondences showing regressive lowering *o > /ɔ/

	'light' *ʤonᶯak	'insect' *poka	'horse' *gᶯoɖa	'shop, store' *d̪okan	'gold' *ʃona	'key' *ʧʰorani	'fat' *moʈa
KS		–	–	d̪ukan [a]	sʌna	–	moʈo [a]
RL	ʤʌnᶯak	pʌka	gᶯʌra	d̪ʌkan	sʌna	ʧʰərani	moʈo [a]
MH	ʤʌnᶯak	pʌka	gᶯʌra	d̪ʌkan	sʌna	ʧʰʌrani	mʌʈʌ
TH	–	pɔka	gᶯɔra	d̪ɔkan	ʃɔna	–	mɔʈa
SH	–	poxa	gᶯoɽa	d̪oxan	ʃona	–	moʈa
RP	–	poka	gᶯoɽa	d̪okan	ʃona	tsʰorani	moʈa
BH	ʤonak	poka	gᶯoɽa	d̪okan	ʃona	tsʰorani	moʈa
BN	Zonak; sonak	poka	kᶯora	dokan; tokan	hona	–	mota

[a] Suspected Hindi loanword.

The change is reconstructed as follows:

[PI 25.] *o > /ɔ/ / _C a {KS, RL, MH, TH}. Diagnostic.[26]

This lowering process is not a common feature of NIA or Magadhan languages—see for example SCB /gʱoɽa/, SCA /gʱora/ 'horse', etc.—and as such, contact with these lects is an unlikely source of the innovation. The change has relatively complex conditioning, being restricted to a specific vowel when it precedes another specific vowel. Given the complexity and the ecological distinctiveness in NIA, this change is unlikely to have developed independently in RL, MH, KS and TH, and is diagnostic of a propagation event between these lects, at some point in their history.

A distinct, morphologically conditioned vowel lowering process is found in southern and eastern KRDS lects, as well as in SCB. The change is as follows: the inherited high vowels *i and *u are lowered one notch in monosyllabic verb roots when followed in the next syllable by a non-high vowel. This change is considered diagnostic of a propagation event on the basis of the complex morphological and phonological conditioning.

[PI 26.] *i, *u > /e/, /o/ / #(C)_C-V[- high] (verb root) {RP, BH, Bangla}. Diagnostic.

For example, /ken-e/ 's/he buys', versus /kin-i/ 'we buy'; /bʱok-e/ 'it barks (as a dog)', /bʱuk-il/ 'it barked'. Of the eight KRDS lects examined, this change only occurs in RP and BH.

Regressive vowel lowering of *i and *u is also found outside the verbal morphology in Rangpur, as the following examples illustrate.

Table 4.37: Example correspondences showing regressive lowering of *i and *u

	'slowly' *ɖ̤ʱirɛ	'thin' *ʧikɔn	'danger' *bipɔɖ̤	'life' *ʤibɔn	'Friday' *ʃukɔr-	'north' *uʈ̤ʈ̤ɔr	'pretty' *ʃunɖ̤ɔr
KS	ɖ̤ʱirɛ	–	–	–	sukər	uʈ̤ʈ̤ər	sunɖ̤ər
RL	ɖ̤ʱirɛ	–	–	ʤiβʌn	sukər-	uʈ̤ʈʌr	sunɖ̤ʌr
MH	ɖ̤ʱirɛ	–	–	ʤiβʌn	sukʌr-	uʈ̤ʈ̤ər	sunɖ̤ər
TH	ɖ̤ʱire	ʧikon	bipɔɖ̤	ʤiβon	ʃukur-	uʈ̤ʈor	ʃunɖ̤ori
SH	ɖ̤ʱire	–	bipɔʈ̤	–	ʃuxur-	uʈ̤ʈor	–
RP	ɖ̤ʱere	tsekon	bepod	ʤebon	ʃokor-	oʈ̤ʈor	ʃonɖ̤or
BH	ɖ̤ʱire	tsiknæi		–	ʃukur-	uʈ̤ʈor	ʃunɖ̤or
BN	–	–	bipɔt	–	hukur-; sukur-	uttɔr	hundɔr

This lowering of *i, *u occurs whenever the following vowel is *ɔ, or *ɛ (though /ɖ̤ʱere/ is the only lexical example in the RP data of a high vowel followed by *ɛ). This phonologically general lowering process is only found in RP out of the eight KRDS lects examined here.

[PI 27.] *i, *u > /e/, /o/ / C_CV$_X$; where V$_X$ = *ɔ, *ɛ {RP}. Diagnostic.

---

[26] *ɔ > ʌ by the phonetic change [PI 29.], see §4.4.5.

As noted above, *o is extremely rare in non-initial syllables, and there are no lexical cases in the data of a high-vowel followed by *o.

There is one instance of lowering of *i in a monosyllabic noun stem: *ḍ̲il > /ḍ̲il; ḍ̲el/ ‘heart’ {RP}. However, this is a sporadic variation currently without broader effect in the lexicon.

4.4.4 Prosodic vowel raising

It has been stated above that *o is rare in non-initial positions. However, the sound [o] has become very common in these positions in some lects through raising of *ɔ > [o]. This change is connected to the initial stress which characterises some KRDS lects, and is shared also with SCB. Klaiman (1990:498) writes:

> Vowel Raising, produces a neutralisation of the high/low distinction in the mid vowels, generally in unstressed syllables. Given the stress pattern of the present standard dialect, … Vowel Raising generally applies in non-word-initial syllables

It is common to define this change for SCB in terms of ‘the mid-vowels’ (cf. also P. Dasgupta 2003), that is, the merger of /ɔ/ and /o/, and /ɛ/ and /e/, in non-initial position. While such a statement may have synchronic value, from a diachronic perspective it is over-defined. The phonemic distinction between /ɛ/ and /e/ was not inherited from Māgadhī Apabhraṁśa, but is an innovation that characterises Bangla, Asamiya, and six of the eight KRDS lects compared in this study (cf. §4.2). In these lects, the phonemes /e/ and /ɛ/ contrast in word-initial syllables only, where /e/ occurs as a result of (i) borrowing, and/or (ii) regressive lowering of /i/ in some lects (cf. §4.4.3), and/or (iii) regressive raising of *ɛ (cf. §4.4.1). To say that /e/ and /ɛ/ are merged in non-initial syllables is effectively to ‘undo’ a phonemic split in a position in which it never occurred. What can be said instead is that (i) the innovative phoneme /e/ is only found in word-initial syllables, and (ii) the inherited vowel *ɛ has developed a lower articulation in word-initial syllables than elsewhere in certain lects—namely SCB and some of the eight KRDS lects (SH, RP, BH, TH).

In contrast to the e/ɛ distinction, the o/ɔ distinction is an inherited feature, that is, both *ɔ and *o are protophonemes. Therefore to describe the raising of *ɔ > /o/ in non-initial position as a loss of phonemic distinction in this position, is historically accurate. As noted above, the raising of non-initial *ɔ is related to the initial stress pattern found in these lects, and is termed ‘prosodic vowel raising’. It is curious that in BN and SH, where there is no strong initial stress, *ɔ has /o/ as an irregular reflex in the *first syllable of the word*. Also, note that in Asamiya ‘Post-accentual *a* [i.e. /ɒ/—MT] is always short … Often, however, this short sound is indicated by *o*’ (Kakati 1962:74–75). The linguistic and historical relations between these phenomena require further analysis, and no conclusion is possible at the present time.

[PI 28.] *ɔ > /o/ / non-initial syllables {RP, TH, BH, variably in SH; also SCB}. Diagnostic of contact with SCB through diglossia.

Prosodic raising of *ɔ > o must have occurred *subsequent* to [PI 23.], the harmonic raising of *o > u; otherwise it would feed that change, which it does not. For example, নিচত *niʧ-ɔṭ̲ ‘under-LOC’ > /niʧ-oṭ̲/ in some lects due to prosodic raising, but this does not feed [PI 23.] to give niʧuṭ̲.

Prosodic raising of *ɔ is found in SCB as well as in the KRDS lects located within the geopolitical scope of Bengal. It seems highly likely that contact through diglossia with SCB has had some role in the spread of this change. As a change propagated through diglossia it would not be diagnostic of a unified propagation event (cf. §3.4.1.2).

4.4.5 Unrounding

The reflexes of *ɔ show a phonetic difference for the western lects KS, RL, and MH as compared with the other more central and eastern KRDS lects.

[PI 29.] *ɔ > /ʌ/ {KS, RL, MH}. Diagnostic of contact relations of diglossia with Hindi.

Of the eight sample lects, these three are the ones (i) located within or near the borders of Bihar, in which the official language is Hindi; and (ii) which also have most evidence of borrowing from Hindi, for example in the nominal postpositions (Ablative, Instrumental, etc. see Chapter 5). As a change plausibly related to diglossia, the range of [PI 29.] is not diagnostic of a unified propagation event (cf. §3.4.1.2 and §7.5.2.2).

While [PI 29.] is a phonetic change, it has phonemic consequences due to the variable articulation of /ʌ/ as [ə] which leads to a merger with [ə] < *a (see §4.4.7).

4.4.6 Transposition and loss of medial high vowels

In several KRDS lects, medial high vowels are either lost or transposed to the position before the preceding consonant. The data in Table 4.38 illustrate the phonological correspondences.

Table 4.38: Example correspondences for transposition and loss of medial high vowels

	_C_XV[+high]C_Ya [a]			
	‘feather’	‘sickle’	‘skin’	‘light’ (not heavy)
	#245	#671	#462	#1284
	*pak^{h}ina	katʃija	*ʧamiɖa	*haɭuka
KS	pək^{h}ina	kətʃia	ʧəmra	həlka
RL	pɛk^{h}na	kɛtʃia	ʧɛmra	hulka
MH	pɛxna	ketʃia	ʧɛmra	həlka
TH	poxina	koitʃa	ʧamra	hulko
SH	paxɛna	kasi	tsameɽa	–
RP	pakna	kaiso	tsamɽa	halka
BH	paxna	katsi	tsamɽa	halka
BN	–	kasi	samra	–

[a] Where either C_X or C_Y is a sonorant consonant.

There are three processes affecting the vowels in this table. Firstly, in some lects and in particular environments, high vowels are transposed to the position before the preceding consonant. This change reduces by one the number of syllables in the word.

[PI 30.] $*V_1C_X\mathbf{V}[+high] > *V_1{}^{\mathbf{V}[+high]}C_X$ / _ C_YV {irregularly in KRDS, Middle Bangla, Oriya of C15th, Kamrupi Asamiya, …}.[27] Diagnostic value unclear (but see §7.4.4 for possible sociohistorical conditioning of propagation).[28]

This change is quite an old change, as shown by its presence in Middle Bangla and Oriya texts. In order to gauge the sociohistorical plausibility of a propagation event involving Middle Bangla, Middle Oriya and KRDS more information is required on the dialectology of this feature. Masica (1991:196) notes the presence of *i transposition in Sadani, a western Magadhan lect. Any conclusions regarding the diagnostic value of [PI 30.] for propagation events need to be based on a Magadhan-wide reconstruction of linguistic history. A broad but interconnected propagation of PI 30 may be connected with the Mughal expansion, but this requires further study (see §7.4.4).

In the second process reconstructed from the data above, a transposed high vowel (termed 'epenthetic' in traditional IA studies) is either deleted, or fused with a preceding *a.

[PI 31.] $*a^i > ɔ, ɛ$ / {TH irregularly}. Nondiagnostic.

[PI 32.] $*a^i > ɛ$ / {RL, MH} (after [PI 30.]). Diagnostic.

[PI 33.] $*V_1{}^{V[+high]} > V_1$ / {RL, MH, SH, RP, BH, and irregularly in TH; also occurs in Middle Bangla and Oriya of C15th} (after [PI 7.]–[PI 8.] and [PI 31.]–[PI 32.]). Nondiagnostic.

[PI 31.] is irregular in TH and cannot be used to diagnose a propagation event. [PI 33.] constitutes a simplification of the complexity of the sequence and (analogous to monophthongisation, see §4.4.9) is not diagnostic of a propagation event because of the possibility of independent and parallel simplification of the sequence.

[PI 32.] is a regular phonological process in RL and MH attested by correspondences in several etyma. The range of propagation across these two lects is sociohistorically plausible based on geographical contiguity. For this reason [PI 32.] is diagnostic of a unified propagation event.

Table 4.38 only includes one example for medial *u in the respective environment, and it is unclear how to interpret the data. RL /hukla/ and TH /hulko/ suggest transposition and replacement of the preceding *a. KS and MH /həlka/, however, confuse the situation unless they constitute loans from Hindi. Further cognates are needed in order to determine this matter for medial transposed *u.

The third process attested by the data in Table 4.38 is the 'shortening' of *a to [ʌ; ə] under certain conditions. This change is examined in the next section.

4.4.7 'Shortening' of *a > [ʌ; ə]

In MH and RL, /ɛ/ is found as a regular reflex of KRDS protophoneme *a, under the conditions outlined in [PI 30.] and [PI 32.]. Under slightly different conditions, KRDS *a gives /ʌ/ in these same lects, as well as in KS. (Recall from §4.2 that the phoneme /ʌ/ in these lects includes subphonemic variation [ə, ʌ]).

---

[27] Either C_X or C_Y is a sonorant consonant. See Chatterji (1926:378ff.) regarding epenthesis in middle Bangla and Oriya.

[28] Chatterji holds that this change is diagnostic of eastern Magadhan.

[PI 34.] *a > /ʌ/ / C_C(C)aC {KS, RL, MH}. Supportive, not diagnostic of contact relations of diglossia with Hindi.

For example, RP [p**a**har] ‘mountain’ compared with KS-RL-MH [p**ə**haɽ] ‘mountain’. The same correspondence is found when two consonants intervene between the two instances of *a, that is: Ca̲CCaC. This change is most regular in KS, with some exceptions present in the data for RL and MH. The change is less regular in RL and MH when the final C is a liquid. In KS, this change also occurs variably without a final C; that is, in the environment C_Ca.

This interchange between *a and [ʌ; ə] is not unique to these lects, but also occurs amongst Hindi and Bihari lects. The development of [PI 34.] in the western-most lects of KRDS is quite likely to have been influenced by contact relations of diglossia with Hindi and therefore [PI 34.] is not diagnostic of a PE in KRDS linguistic history. There is a structurally similar change in Asamiya also, *a > ɒ / _Ca, but it is almost certainly a separate innovation to [PI 34.].

4.4.8 Vowel insertion between consonant clusters

In an almost opposite process to that analysed in §4.4.6 above (medial high vowel deletion), in Shalkumar a vowel is regularly inserted between consonant clusters if the second consonant of the cluster is either /l/, /n/, or /r/, for example *hamra > /hamɛra/ ‘we’. The inserted vowel is regularly conditioned by features of the vowel in the preceding syllable. The relevant features are +/-high, +/-back.

Table 4.39: Example words showing vocalic insertion in SH

	+high -back	+high +back	-high -back	-high -back	-high +back	-high +back
Preceding vowel:	*i	*u	*ɛ	*a	*o	*ɔ
Inserted vowel:	/i/	/u/	/ɛ/	/ɛ/	/o/	/o/
Gloss	‘they:PROX’	‘they:DIST’	‘crooked’	‘feather’	‘you: NOM PL’	‘blunt’
Et. Id.	#2289	#2290	#1333	#245	#2288	#1294
pKamta	*[ɛmⁿra; imⁿra]	*[omⁿra; umⁿra]	*bⁿɛkɖa	*pakʰina	*t̪omⁿra	*bⁿɔt̪ra
SH	/imira/	/umura/	/bⁿɛxɛra/	/paxɛna/	/t̪omora/	/bⁿɔt̪ora/
Gloss	‘inside’		‘mouse’	‘bullock’		‘question’
Et. Id	#1011		#300	#263		
pKamta	*bⁿiṯɔr-[a]		*nɛknaj	*ɖ̪amura > *ɖ̪amra		< prɔʃnɔ
SH	/bⁿiṯira/		/nɛxɛnai/	/ɖ̪amɛra/		/pɔsonɔ/

[a] *bⁿiṯɔr- {PKmt}> *bⁿiṯɔra > *bⁿiṯra > /bⁿiṯira/ {SH}

The change is analysed as follows:

[PI 35.] *Ø > V[α back, β high] / V[α back, β high] C_C_XV, where C_X = /l/, /n/, or /r/ {SH}. (After [PI 30.] and [PI 33.]). Diagnostic.

As the examples show, this change is fed by other changes which establish the conditioning environment. The presence of medial /ɛ/ in SH /paxena/ is not best explained as a mutation of the inherited *i in *pak^{h}ina, as we find medial *i with reflex /i/ in other items, for example /d̪orina/ 'river bank' < *d̪ɔrina. The /ɛ/ in SH /paxena/ is instead best explained by reconstructing medial transposition and deletion of *i ([PI 30.] and [PI 33.]) followed by the medial insertion rule [PI 35.]. The conditioning of this insertion is complex, and as a result the change is diagnostic of a propagation event.

4.4.9 Inherited diphthongs

Inherited diphthongs are treated in this study as a sequence of one of the protovowels followed by approximants *j or *w. Such sequences reconstructed in proto Kamta vocabulary are usually retained without change in the present-day lects, though an exception requires some further comment here as it affects the reconstruction of verbal morphology.

The exceptional case is *ɔw. This reconstructed protosequence is justified by the following comparable Tadbhava forms. The only irregularity is in the cognate set for *kɔwn- 'where'. In some cases this irregularity is the result of raising due to a high vowel in the pronominal base, for example BN /kuti/ 'where'; in other cases it is not explained, e.g. SH /kunʈa/ 'where'.

Table 4.40: Cognate forms exhibiting *ɔw

Gloss	'straw'	'iron'	'black-smith'	'where?'	'when'	1SG in AGR.I
Et. Id.	#58	#99	#816	#1230, #1233	#1214	
pKmt	*pɔwalʂ	*lɔwha	*lɔwharɔ	*kɔwn-	*kɔwn-	*-ɔw˜
KS	pwal; pol	lua	luhar	kun^{ɦ}a	–	-u
RL	pual	luha	luhar	kun^{ɦ}a; kun-ʈ^{h}a; kun-ʈ^{h}ina	kun-bɛla	-u
MH	pual-i	luha	–	kunʈ^{h}ɛ; kunʈ^{h}in	kun-bɛla	-u
TH	pwal	luha	–	kunʈ^{h}e	kunbɛla	-u
SH	poal	noha	–	koʈe; kun-ʈa	–	-o
RP	poal	nowa	–	koʈe	konbæla	-o˜
BH	poal	noha; noa	–	koʈe; kuʈi	konbela	-oŋ
BN	–	loha	–	kuti	konbela; kunbela	-o˜

Reconstructing the morpheme in the rightmost column as *-ɔw˜ also fits well with attested late MIA forms. The Apabhaṁśa primary ending for first person singular is *-auṁ*, alternatively Romanised as *-au˜* (Bubenik 2003, see Table 6.16) where *-a* represents the short central vowel from which we get eatern Magadhan /ɔ/. This reconstruction involves the following changes:

[PI 36.] *ɔw > /u/ {KS, RL, MH, TH}. Supportive, not diagnostic.

[PI 37.] *ɔw > /o/ {SH, RP, BH, BN }. Supportive, not diagnostic.

The diagnostic value is at best suggestive or supportive of a propagation event because the monophthongisation of this diphthong is far from unique in NIA.

4.4.10 Vowel nasalisation

In some KRDS lects, vowels are distinguished as +/- nasalisation—a phonological feature inherited from MIA. The retention of this feature is not uniform across KRDS, and is also conditioned in certain lects by the position of the nasalised vowel in a word. Though certain generalisations are possible, there are a few exceptions to these rules. In Bongaigaon, nasalisation has generally been lost in initial syllables, while in Kishanganj and Rangeli it has been retained in initial syllables. In Rangpur, the nasalisation is retained only in monosyllabic words and in verbal suffixes (cf. §6.3). In the other lects, nasalisation is generally lost. These patterns of correspondence are illustrated in Table 4.41, with dark shading indicating the loss of nasalisation. Nasalisation of a vowel is marked by a tilde immediately to the right of the vowel, as in Masica (1991:xvi).

Table 4.41: Examples of inherited nasalisation, in different positions

	monosyllabic [a]		disyllabic, 1st σ	disyllabic, 2nd σ
Gloss	'camel'	'grass'	'thorn'	'mud'
Etym.	*uṣṭra-*	*ghāsa-*	*kaṇṭaka-*	*kardama-*
Et. Id.	#293	#51	#48	#117
pKmt	*u˜ʈʰɔ	*gʱa˜ʃɔ	*ka˜ʈa	*kaɖɔ˜
KS	u˜ʈ	gʱa˜s	ka˜ʈʌ	kaɖo
RL	u˜ʈʰ	gʱa˜s	ka˜ʈa	kaɖʌ
MH	uʈʰ	gʱas	kaʈə	kaɖʌ
TH	uʈʰ	gʱaʃ	kaʈa	kaɖa
SH	uʈ	gʱaʃ	kaʈa	kaɖo
RP	u˜ʈ	gʱa˜ʃ	kaʈa	kaɖo
BH	uʈ	kʱaʃ	kaʈa	kaɖo
BN	utʰ	kʱas	kata	kadɔ˜

[a] After the loss of final *ɔ by [PI 38.].

Nasalisation is generally retained in the neighbouring related lects SCA, SCB and SCO, as well as in Hindi, Nepali and Bihari. The neighbouring presence of nasalisation contrasts with its loss in some KRDS lects. This contrast suggests that the range of the loss should be explained with reference to propagation events. However, there are four distinct patterns to the loss of nasalisation in KRDS: BN versus RP versus KS-RL vs. the rest. Furthermore, there are several logical possibilities for how these four patterns came about. For example, nasalisation could have been lost in noninitial syllables in all but BN, with the other three patterns subsequent to this change. Alternatively, nasalisation may have been lost firstly in nonmonosyllabic words in all lects but RP, with the other patterns a later development, etc. etc.

Unlike in the case of the palatalisation versus labialisation in §4.3.3, there is no clear linguistic diachronic principle which provides a natural ordering to the loss of nasalisation in the three specific environments: first syllable, other syllable, monosyllable. The best candidate for a guiding linguistic principle may relate to the loss of nasalisation in non-stressed position. However, testing this hypothesis requires closer synchronic study of the conditioning of stress in KRDS lects than has been undertaken to date. At present therefore, loss of nasalisation cannot be used to reconstruct propagation events in KRDS history.

4.4.11 Treatment of inherited final vowels

In §4.3.9 it was argued that two consonant phonemes—postalveolar (retroflex) nasal and equivalent lateral—must be included within the proto Kamta phoneme system in order that it be a realistic reconstruction of the probable consonant system of proto Kamta. The loss of these two phonemes throughout KRDS, as well as much of NIA, does not alter the reconstruction—attested by textual and comparative evidence—that they were in all likelihood present in KRDS lects up to at least the 15th century, as in Bangla and Maithili.

An analogous situation pertains for the reconstruction of final vowels in KRDS linguistic history. Some inherited final vowels have been recently lost in KRDS, as is the case for most NIA lects. However, textual evidence from earlier NIA, not to mention MIA, clearly shows them to have been present during these earlier stages. Furthermore, regular retention of these final vowels still persists in a few unconnected NIA lects—including Sindhi, Maithili and Oriya. Chatterji (1926:301) analyses the chronology of this change based on Middle Bengali literature:

> Final vowels of OIA. were continued down to Late MIA. and Early NIA. times. The long final vowels «-ā, -ī, -ū», however, were shortened to «-ă, -ĭ, -ŭ», and «-ē, -ō» were weakened and shortened to «-ĭ, -ŭ» in late MIA. (Apabhraṁśa), and these shortened vowels «-ă, -ĭ, -ŭ» fell together with original short «-ă, -ĭ, -ŭ». All NIA. inherited these short vowels, but in later times in Bengali, during the Middle Bengali period, they were all dropped, or assimilated: except «-ĭ, -ŭ» where they were preceded [with no intervening consonant—MT] by vowels of a different quality.

The chronology proposed by Chatterji for the loss of final vowels (Middle Bengali period) is similar to that proposed for the merger of dental and postalveolar nasals and laterals (cf. §4.3.7). These changes are also alike in being widely distributed across NIA lects (see §7.4.4). Note that in Indic scripts a final sequence of *Cɔ is only distinguishable from a final sequence *C# if the syllable terminating diacritic (called virāma, halant, or hôshonto) is used consistently to indicate consonant final syllables. In NIA, it is questionable whether the use of the halant is sufficiently consistent to justify this assumption. It is thus difficult to determine, on textual criteria, the chronology of the loss of final *ɔ.

The pattern in Asamiya with respect to loss of final vowels is generally the same as for Bangla, with the exception of inherited final *i. In the case of some Asamiya words this vowel has been lost in word final position, for example (from Kakati 1962:94): /tɒrowal/ < *taravāri* ‘sword’; /gabʱin/ < *garbhinī* ‘big with young’.[29] In the case of other Asamiya words, the final *i is retained, for example (ibid.): /rati/ < *rātri* ‘night’; /sari/ < **cattāri-*; *catvāri-*

[29] Oddly, Kakati also gives the form /gabʱini/ < garbhinī ‘big with young’ as evidence for the retention of inherited final *i.

'four'. Kakati hypothesises that in the case of the latter set of words the final *i was retained due to suffixation with pleonastic [*-ka*; *-kā*]. (These segments would have then been eroded during early NIA through lenition of intervocalic *k* and loss of the final vowel [*-a*; *-ā*]). A different explanation is given by Masica (1991:196) for the variable retention of final *i in Asamiya, which he links to similar reflexes in a western Magadhan lect:

> final *-i* is 'preserved' in [Sadani] by being in effect transposed to the preceding syllable: *āig* 'fire' (< **āg'* < MIA *aggi*), *dāil* 'split pulse' (< MIA *dāli*), *rāit* 'night' (< MIA *rattī*). This occurs occasionally elsewhere in NIA (most frequently in Assamese).

Transposition of this kind (which Chatterji and Kakati term 'epenthesis') 'while … not a noticeable feature in the standard colloquial [i.e. eastern Asamiya—MT], … is a distinguishing characteristic of Western Assam colloquy' (Kakati 1962:148). It is possible that this difference between eastern and western Asamiya lects with regard to /i/ transposition lies behind the variable retention of final /i/ in Asamiya. In the history of western lects, final *i was transposed to before the preceding consonant. In this position the *i would not be deleted, as it is no longer in final position. Contact between speakers of western and eastern Asamiya plausibly led to the sporadic reintroduction of final *i in some words through copying the *i element *which was maintained in Western Asamiya by transposition*. A precise reconstruction of this hypothetical process requires close knowledge of the historical Asamiya literature, and is beyond the scope of this study.

The regular reflexes of inherited final vowels in KRDS (shown in Table 4.42) are similar to Asamiya in that final *i is retained, but in KRDS final *a and *u are also frequently retained. KRDS final *a does not go back to OIA final *-ā* –that etymological phoneme merged with etymological final *-a* (> KRDS *ɔ) during a late MIA process (see quote from Chatterji above). Hence, there are no Tadbhava words retained in KRDS with a final *-a traceable to OIA final *-ā*. Nonetheless there are regular reflexes of *a in final position that are well distributed across KRDS. The distribution across KRDS of these forms with final *-a is unlikely to be the result of recent propagation—the appropriate sociohistorical conditions for such a propagation have not been present for almost five hundred years—but are more likely to have been inherited from the common proto Kamta period of linguistic history. Occurences of final *-a in KRDS are either

- the result of innovative affixation of a nominal suffix *-a, for example *ʤibⁿa 'tongue';
- a Tatsama (Sanskrit loanword), for example *indra 'well', or:
- a Persian loanword, for example *hawa 'wind, breeze, air'.

The function of the nominal suffix *-a mentioned above is described by Chatterji (1926:348–349) as follows:

> the «-a» in the form কালা «kala» is an affix giving a definite force,=*the black one* ; and this can only be from some affix like «-ā-ka», with a definiteness that came to be associated with «-ā» … The «-ā» nouns and verbal adjectives of Bengali and other NIA. are probably to be referred to oblique (genitive) forms of Late MIA. and Early NIA. Where the original nominative affixes were lost, in some forms of NIA., it is this oblique in «-ā» that took its place.

The suffixation of particular nouns with this morpheme *-a is innovative, and possibly could be diagnostic of propagation events. However, as this suffixation process is found across NIA, the reconstruction of the propagation events involved is beyond the scope of this study.

With the exception of some morphological suffixes, the phonemes *o and *ɛ are not reconstructed in word final position. This is largely because in final position OIA *-e* and *-o* merged during late MIA (Apabhraṁśa) with OIA *-i/ī* and *-u/ū* respectively (see the quote from Chatterji above). Shaded cells in Table 4.42 are proposed in this reconstruction to be borrowings from Hindi (in the case of KS and RL) or Bangla (in the case of TH and RP).

Table 4.42: KRDS reflexes of proto Kamta final *i

Gloss		'stick'	'tama rind'	'night	'wall'	'rain'
Et. Id.		#16	#70	#1495	#617	#149
pKmt	*-i	*laʈʰi	*t̪ɛt̪ɔli	*rat̪i	*ʈaʈi	*ʤʱɔɖi
KS	i	laʈʰi	t̪ɛt̪ul [a]	rat̪	ʈaʈi	–
RL	i	laʈʰi	t̪ɛt̪li	rat̪	ʈaʈi	ʤʱʌri
MH	i	laʈʰi	t̪it̪li; t̪ɛt̪li	rat̪i	ʈaʈi	–
TH	i	laʈʰi	t̪et̪ul	rait̪; ait	ʈaʈi	–
SH	i	naʈʰi	t̪it̪ili	at̪i	ʈaʈi	zʱoɽi
RP	i	naʈi	t̪eit̪ol	ait̪	ʈaʈi; ʈaɽi	ʤʱori
BH	i	naʈi	t̪et̪li; t̪et̪ili	rat̪i; nat̪i	–	ʤʱori; tsʱori
BN	i	latʰi	titili	rati	–	–

[a] Bengali loanword.

Table 4.43: KRDS reflexes of proto Kamta final *u

Gloss		'palate'	'eye'	'thin'	'some'
Et. Id.		#446	#365	#1315	#1198
pKmt	*-u	*t̪aɭu	*ʧɔku	*ʃoru	*kiʧʰu
KS	u	t̪alu	ʧouk	–	kuʧʰu
RL	u	t̪al-ka	ʧʌkʰu	suru	kiʧʰu
MH	u	t̪al-ka	ʧoxu	suru	kiʧʰu
TH	u	t̪alu	ʧok	–	kiʧʰu
SH	u	t̪alu	tsoxu	ʃoru; ʃoɽu	kesu
RP	u	t̪alu	tsouk	ʃoru	–
BH	u	t̪alu	tsoxu	ʃoru; ʃoɽu	kisu
BN	u	talu	suku	horu	–

In some words listed in Tables 4.42–4.43, the high vowel has been transposed to before the preceding consonant. This process is analysed in §4.4.6 and need not detain us here.

Table 4.44: KRDS reflexes of proto Kamta final *ɔ

Gloss		'cold'	'potter'	'farmer'	'bed'
Et. Id.		#1267	#828	#826	#692
pKmt	*-ɔ	*ʤaɖɔ	*kumᶯarɔ	*kiʃaɳɔ	*kʰaʈɔ
KS	/	–	kumᶯar	kisan	kʰaʈ
RL	/	–	kumᶯar	kisan	–
MH	/	ʤar	kumᶯar	kisan	–
TH	/	ʤar	kumᶯar	kiʃan	–
SH	/	–	kumar	kiʃan	kʰaʈ
RP	/	ʣar; ʣaʈ	kumær	–	kʰaʈ
BH	/	ʣar	kumær	kiʃʲæn	kʰaʈ
BN	/	–	kumar	–	kʰat

Table 4.45: KRDS reflexes of proto Kamta final low vowels

Gloss		'tongue'	'iron'	'insect'	'kite'
Et. Id.		#468	#99	#246	#327
pKmt	*-a	*ʤibᶯa	*lɔwha	*poka	*ʧilᶯa
KS	-a	ʤiβa	lua	poka	ʧilᶯa
RL	-a	ʤiβᶯa	luha	pʌka	ʧilᶯa
MH	-a	ʤibᶯa	luha	pʌka	ʧilᶯa
TH	-a	ʤiβᶯa	luha	pɔka	ʧilᶯa
SH	-a	ziβᶯa	noha	poxa	tsiɕa
RP	-a	ʣibʲæ	nowa	poka	tsiɕæ
BH	-a	ʣibʲæ	noha; noa	poka; poɣa	tsilʲæ
BN	-a	ziβᶯa; siβᶯa	loha	poka	sila

The etyma in Table 4.44 show the loss of a hypothetical final *-ɔ. The evidence from Oriya suggests that final *-ɔ was inherited as part of the common Magadhan stage, derived from OIA final *-a* and *-ā*. Final *ɔ was then lost subsequent to the breakup of proto Magadhan. Further chronology is presently uncertain.

[PI 38.] *ɔ lost word finally {KRDS, SCB, SCA, Hindi, Bhojpuri etc.} (chronology uncertain)

As the chronology of this loss is presently unclear, the proto Kamta vocabulary has been reconstructed to *include* the final *ɔ though this reconstruction may need to be revised if it can be shown that [PI 38.] occurred prior to AD 1550 when the proto Kamta speech community is reconstructed as having undergone division (see §7.3.1).

It turns out that of the six inherited proto Kamta vowel phonemes, only *ɔ is lost word-finally, by [PI 38.].

4.4.12 Summary of reconstructed vowel system

Based on the arguments presented in this chapter, KRDS is reconstructed as having inherited six segmental vowel phonemes from the proto Kamta stage (cf. §7.3.1) after its split from proto Magadhan or some intermediary protolanguage.

Table 4.46: Proto Kamta vowel system, inherited by KRDS lects

	Front, unrounded	Back, rounded
High	*i	*u
Mid	*ɛ	*o
Low	*a	*ɔ

References to the relevant sections dealing with the changes to protovowels are given in Tables 4.31 through 4.34 and are not repeated here. The inherited KRDS vowel system has no phonemic length contrast—a feature whose loss is shared with other Magadhan lects, and inherited from a proto Magadhan stage. Vowel nasalisation has been reconstructed as an inherited feature of KRDS, despite the fact that individual changes to nasalisation are not recoverable at present.

4.5 Summary of diagnostic phonological innovations

The following changes are found to be either diagnostic of propagation events in the linguistic history of KRDS or supportive of PEs diagnosed by other changes:

[PI 1.] Devoicing of the obstruent element (not the aspiration) of initial voiced aspirates {regular in BN, variable in BH}. Diagnostic.

[PI 3.] *d̤, *ɖ̤, *ʤ, *b^{ɦ}, *g^{ɦ} > [-voice] / _# {SH, BH, BN} (after rhoticisation). Supportive, not diagnostic.

[PI 6.] *C > C^{j} / i _ a {SH, RP, BH} (co-occurent with [PI 7.]). Diagnostic.

[PI 7.] *C > *C^{w} / u _ a {SH, RP, BH} (co-occurent with [PI 6.]). Diagnostic.

[PI 8.] *C^{w} > C^{j} {RP, BH} (after [PI 7.]). Diagnostic.

[PI 9.] *C[+breathy voice, +continuant] > [+modal voice] / V_V {SH, RP, BH, BN, Oriya, SCA, SCB} ([tentatively] after C16th, after rhoticisation). Diagnostic.

[PI 10.] Loss of aspiration in all intervocalic consonants {RP, BH, ?SCB, ?Oriya}. Diagnostic.

[PI 11.] Apical series > alveolar articulation {BN and SCA} (during or before C15th in Asamiya, C20th in BN). Diagnostic of contact relations with SCA through diglossia.

[PI 12.] *ɳ , *ɭ > /n, l/ {KRDS, SCB, SCA, Maithili, Hindi, etc.} (C15th or later). Possibly diagnostic based on sociohistorical plausibility (cf. §7.4.4) but broader NIA reconstruction is necessary in order to verify the diagnostic value of this change..

[PI 14.] *l > /n/ / #_ {RP, SH, BH, and TH Hindus}. Supportive, not diagnostic.

[PI 15.] *r > Ø / #_ {RP, variably in SH, BH and among TH Hindus}. Diagnostic

[PI 16.] *r > n / #_ {BH variably}. Diagnostic.

[PI 17.] Homorganic cluster of N C [+asp, +vc] > N [+asp] {irregularly in KS, RL, MH, TH} Supportive, not diagnostic.

[PI 18.] *ɛnḍurɔ > /niḍur/ 'rat' {KS, RL, MH, TH}. Supportive, not diagnostic.

[PI 19.] *ʃ > /s/ {KS, RL, MH (from Hindi, Nepali)}. Supportive, not diagnostic, of contact through diglossia with Hindi and/or Nepali.

[PI 22.] *ɔ > [o] / _ (C) V[+high] {TH, SH, RP, BH, BN, SCB and SCA}. Supportive, not diagnostic.

[PI 23.] *o > /u/ / V[+high]C_ {RP, BH}. Diagnostic

[PI 24.] *a > /æ; ɛ/ / V[+high]C_ {RP, BH}. Diagnostic.

[PI 25.] *o > /ɔ/ / _Ca {KS, RL, MH, TH}. Diagnostic.

[PI 26.] *i, *u > /e/, /o/ / #(C)_C-V[-high] (verb root) {RP, BH, SCB}. Diagnostic.

[PI 27.] *i, *u > /e/, /o/ / C_CV$_X$; where V$_X$ = *ɔ, *ɛ {RP}. Diagnostic.

[PI 28.] *ɔ > /o/ / non-initial syllables {RP, TH, BH, variably in SH; also SCB}. Diagnostic of contact with SCB through diglossia.

[PI 29.] *ɔ > /ʌ/ {KS, RL, MH}. Diagnostic of contact relations of diglossia with Hindi.

[PI 30.] *V$_1$C$_X$$^{V[+high]}$ > *V$_1$$^{V[+high]}C_X$ / _ C$_Y$V {irregularly in KRDS, Middle Bangla and Oriya of C15th, Kamrupi Asamiya, …} Diagnostic value unclear (but see §7.4.4 for possible sociohistorical conditioning of propagation).

[PI 32.] *a^i > ɛ / {RL, MH} (after [PI 30.]). Diagnostic.

[PI 34.] *a > /ʌ/ / C_C(C)aC {KS, RL, MH}. Supportive, not diagnostic of contact relations of diglossia with Hindi.

[PI 35.] *Ø > V[α back, β high] / V[α back, β high] C_C$_X$V, where C$_X$ = /l/, /n/, or /r/ {SH}. (After [PI 30.] and [PI 33.]). Diagnostic.

[PI 36.] *ɔw > /u/ {KS, RL, MH, TH}. Supportive, not diagnostic.

[PI 37.] *ɔw > /o/ {SH, RP, BH, BN }. Supportive, not diagnostic.

[PI 38.] *ɔ lost word finally {KRDS, SCB, SCA, Hindi, Bhojpuri, etc.} (chronology uncertain)

The sociohistorical conditioning of the propagation of these changes is examined in Chapter 7. The next two chapters use the phonological reconstuction of this chapter to inform reconstruction of inherited morphemes and the changes which have led to the present-day nominal and verbal morphology.

5 *Reconstruction of nominal morphology*

5.1 Introduction

With this chapter the reconstruction moves from linguistic changes which are phonologically general, to changes that specify the form and function of nominal inflections. Three categories of inflections are central to nominal morphology in KRDS: case markers, specificity classifiers, and pronominals. Case markers (§5.3) indicate the grammatical function of a noun phrase (NP), while specificity-classifiers (§5.4) indicate its discourse function (hence 'specificity'), grammatical class (hence 'classifier'), and also number.

In KRDs, as in Indo-Aryan generally, there are personal pronominals (e.g. 'she') and non-personal pronominals (e.g. 'this much,' 'here'). These are defined by paradigmatic relations between a proximal form (which begins with *ɛ), a distal form (which begins with *o), an interrogative form (which begins with *k-), and a relative form (which begins with *ʤ-). An example of non-personal pronominals taken from Mahayespur:

- /ɛt̪ɛla/ 'this many'—the proximal form;
- /ʌt̪ɛla/ 'that many'—the distal form;
- /kɛt̪ɛla/ 'how many?'—the interrogative form;
- /ʤɛt̪ɛla/ 'as many'—the relative form (i.e. a subordinating conjunction).

The KRDS systems of non-personal pronominals are described, and the proto Kamta system reconstructed, in §5.7.

These four morphosyntactic categories also apply to the personal pronominals which in addition are marked for the grammatical function of the NP in the clause (either as a nominative or oblique argument). For example (again from Mahayespur):

- /ɛɛ/ 's/he here'—proximal—with oblique counterpart /ɛ-/;
- /ʌɛ/ 's/he there'—distal—with oblique counterpart /ʌ-/;
- /kaɛ/ 'who?'—interrogative—with oblique counterpart /kaha-/;
- /ʤahɛ/ 'who' (subordinating conjunction), with oblique counterpart /ʤaha-/.

The personal pronouns for each of the eight KRDS lects are described in §5.5 and the proto Kamta pronouns are reconstructed in §5.6.

The inflectional category of number is covered in the section on specificity-classifiers (§5.4). In KRDS, as in the other eastern Magadhan lects (Oriya, Bangla, Asamiya, etc.), gender is not an inflectional category and hence does not figure in this reconstruction.[1]

These three categories of morphemes—case, classifiers, pronominals—are termed *inflectional* under the broad definition adopted by Masica (1991:212ff., following Zograph 1976), which includes both agglutinative and certain analytic elements 'entering into *paradigmatic contrasts*'. The reason given by Masica for adopting this approach is the non-discrete line between analytical elements and agglutinative affixes—'the former generally ancestral to the latter, through gradual phonetic reduction and adhesion to the stem' (ibid.).

The results of this chapter are a reconstruction of inheritance and change in the inflectional nominal morphology from proto Kamta down to the eight sample KRDS lects. Reconstructed innovations are scrutinised so as to diagnose propagation events (cf. §3.4.1).

It is not the purpose of this study to pursue exhaustively the MIA (Middle Indo-Aryan) and OIA (Old Indo-Aryan) etymologies of the inherited forms. Etymologies are included only as is necessary for distinguishing inherited and innovative features. Closer reference to the ancient morphological systems is not required because of the general discontinuity between OIA and NIA morphology—with MIA morphology transitional between the two. See Chatterji's (1926:715) comments in the context of Bengali historical morphology:

> Bengali like most NIA. languages may be said to have started *de novo* in its morphology, having preserved but very little of the declinational system of OIA.; and the little that it has preserved consists of a few inflexions which have been generalised. MIA considerably curtailed the elaborate declension of the noun of OIA.

The sources of morphological data for non-KRDS lects are:

- Oriya (Dash 1982; Misra 1975; Ramachandran 2001);
- SCA (Goswami and Tamuli 2003; Kakati 1962);
- Rajshahi Bangla (Chaudhuri 1940; Islam 1992);
- SCB (Chatterji 1926, P. Dasgupta 2003);
- Kamrupi Asamiya (U. Goswami 1970);
- Hajong (Haldar 1986);
- Bhojpuri (Shukla 2001; Tiwari 1960);
- Maithili (Jha 1985 [1958]);
- (Dangaura) Tharu (Boehm 2004, pers. comm.).

In addition, the following sources provide data on several lects:

- D. Dasgupta (1978) for Kharia Thar, Lodha, Mal-Pahariya and Manbhum Pahariya;
- Masica (1991) for Awadhi, Bhojpuri, Maithili, Oriya, SCA, SCB and SCH.
- Purkhait (1989) for non-standard Asamiya, Bangla and Oriya geographical 'dialects'.

---

[1] The presence of gender in inflectional morphology for wMg (Bihari) suggests that the loss of this inflectional category is diagnostic of the *eMg stage. However, establishing this conclusively is beyond the scope of the present reconstruction.

5.2 NP structure in KRDS

The Noun Phrase in KRDS contains a head plus optional modifiers that precede the head, such as demonstratives, possessive phrases, quantifiers, and adjectives.[2] The head noun is followed by specificity-classifiers and then case markers: Noun–(specificity-classifier)–(case)

Speaking generally for eastern Magadhan, including KRDS, the structure of NP constituents is as follows.

- *NP constituents precede the head*, with the exception that in some lects (such as BH) a numeral may occur in either of two positions. Firstly, the numeral may occur before an adjective and the head, for example /tin-ʈa kala goru/ 'three-CLF black cows'. Secondly, the numeral may occur after the head, for example /kala goru tin-ʈa/ 'black cows three-CLF' = 'the three black cows'. In lects that allow the second position, the numeral's position with respect to the noun differentiates *specific non-definite* from *definite* pragmatic inferences (e.g. 'three of the cows' vs. 'the three cows').[3] These two positions for numerals with these pragmatic functions are also found in Bangla (P. Dasgupta 2003:379ff.) and Asamiya (Goswami and Tamuli 2003:433ff.).
- Possessive Phrases (PossPhr) precede all other NP constituents;
- Demonstratives precede other modifiers.

Based on these three general points, the structure of the NP (after some simplification, especially concerning CLF position) is:

NP → PossPhr, DEM, NUM, ADJ, N-CLF-Case or:

PossPhr, DEM, ADJ, N, NUM-CLF-Case

Classifiers occur in one of three positions: suffixed to quantifiers, nouns, or demonstratives. When a quantifier is present in the NP, then the post-quantifier position is mandatory for classifiers.

With the exception of classifier marking on quantifiers and occasionally on demonstratives, NP dependents are not inflected for any grammatical features of the head noun.

5.3 Case and case-like postpositions: description and reconstruction

5.3.1 Synchronic overview

Case markers establish the function of the noun phrase within the clause. NPs may be core or adjunct arguments to the verb—the core arguments being the S of intransitive clauses, and the A and O of transitive clauses (Andrews 1985). In KRDS, the S and A arguments are unmarked. The O argument is marked with the Dative-Accusative (DAT) suffix if the head noun is either [human], or [animate and discourse prominent], and otherwise unmarked (cf. §5.3.6). Adjunctive arguments are marked by:

---

[2] See Wilde (2008:287–290) for a more detailed description of noun phrase structure in central Jhapa Rajbanshi.

[3] The notion of specificity entails that the speaker intends for the NP to refer to a unique entity in the world. Specificity is thus different from definiteness, in that definiteness requires that the addressee also be able to uniquely identify the NP referent(s).

- Inflectional suffixes, which are phonologically bound to the noun (or its classifier);
- Postpositions, which occur after a noun in Genitive case;
- Postpositions which, while not grammatically bound to the noun, occur directly after it without an intervening Genitive marker.

In KRDS, the Dative-Accusative (DAT), Genitive (GEN), and Locative (LOC) case markers fit into the first category of inflectional suffixes. For example (from Bhatibari):

- /manʃi-ʈa/ 'man-CLF' = 'the man', unmarked for case;
- /manʃi-ʈa-**k**/ 'man-CLF-**DAT'** = 'to the man', marked for dative-accusative case;
- /manʃi-ʈa-**r**/ 'man-CLF-**GEN'** = 'of the man', marked for genitive case.
- /manʃi-ʈa-**ʈ̪**/ 'man-CLF-**LOC'** = 'in the man', marked for locative case.

The Instrumntal (INS) and Ablative (ABL) postpositions occur directly after the (optional) specificity-classifier, directect after the head noun.For example (again from Bhatibari):

- /laʈʰi **ɖ̪ia**/ 'stick **INS'** = 'with a stick'
- /gatʃ-kʰan **ʈ̪ʰaki**/ 'tree-CLF **ABL'** = 'from the tree'

The position of general postpositions is after a Genitive marked noun phrase. For example:

- /gatʃ-**ɛr ʈ̪ɔl-ɔʈ̪**/ 'tree-**GEN under-LOC**' = 'underneath the tree'

In this example the postposition /ʈ̪ɔlɔʈ̪/ 'underneath' is external to the NP, as shown by the genitive case marker that intervenes between the head of the NP /gatʃ/ 'tree' and the postposed word /ʈ̪ɔlɔʈ̪/ 'underneath'.

The morphological elements included in the reconstruction of this chapter generally exclude the broader category of postpositions (e.g. /ʈ̪ɔlɔʈ̪/ 'under'), in favour of morphology that is internal to the noun phrase (as indicated by the syntactic position: right of the noun head with no intervening case marker). For postpositions other than the Instrumental and Ablative markers, the Genitive case marking indicates that they have not been grammaticalised within the noun phrase.[4]

The case-marking words and suffixes identified here fit into Masica's Layer II of Indo-Aryan Case markers, defined as:

> (a) attached to the base indirectly, through the mediation of a Layer I element; and/or
> (b) invariant for all nouns and the same for both numbers. (1991:232)

The difference between Layer I and Layer II in Masica's scheme is that Layer I elements 'attach directly to the base, with morphophonemic adjustments which are occasionally complex' … 'Morphophonemic variation, while not entirely absent at Layer II, tends to be of a simpler order than in Layer I' (ibid. 231–232). Layer I morphemes, under Masica's definition, are entirely absent from KRDS as for Asamiya. Bangla has the general oblique marker /-ɖ̪/ in the case of plural nouns, and Oriya has /-[ɔ]ŋ/ with the same function. Otherwise NP functions in eastern Magadhan are marked exclusively by elements belonging to Masica's Layer II and Layer III.

---

[4] This syntactic distinction could also be tested for adhesion and the insertability of adverbial or intensifier elements before the postposition. Such testing is outside the scope of the present study.

The elements described above as 'postpositions external to the noun phrase' constitute Layer III of case marking in Masica's scheme. In addition to the syntactic criterion, he also gives a semantic criterion for distinguishing Layer II from Layer III elements:

> [A Layer III element] is semantically more specific. E.g., as compared with a more diffuse Locative on Layer II or perhaps Layer I, Layer III typically mediates such concepts as 'on top of', 'under', behind', 'inside of', 'near', etc. (1991:235)

The Comparative (CMP) marker in KRDS is certainly a Layer III element, rather than Layer II, and on that basis should be excluded from this section. However, it is included in the reconstruction as a special case because of its close functional similarity with the ABL marker. The ABL and CMP markers are addressed jointly in §5.3.9.

5.3.2 Oblique argument marking in eMg

The term 'oblique' is used in Indo-Aryan studies (and the convention will be followed here) to refer to *any case-marked NP*. Thus 'oblique' in this context refers not only to NPs in adjunctive arguments of a clause, but also to NPs with O function, or even A function, if they are marked with an overt case marker.

In KRDS there is no general marker for oblique arguments. An 'oblique' ending (to which case markers attach) exists only in the pronouns, and even the pronominal declension is missing for some KRDS lects (§5.5 and §5.6.1). This is an important point of difference with the Bangla system of nominal declension. In Bangla, 'oblique' (i.e. case-marked) nouns are suffixed first by /-ḍ̲-/ then by the case marker. This oblique declension is limited to semantically animate nouns. For example, in Bangla:

- /manuʃ-ʈi/ 'man-CLF' = 'the man';
- /manuʃ-era/ 'man-NOM.PL' = 'the men';
- /manuʃ-ʈi-r baʃa/ 'man-CLF-GEN home' = 'the man's home'
- /manuʃ-ḍ̲-er baʃa/ 'man-OBL.PL.AN-GEN home' = 'the men's home'

In KRDS, on the other hand, there is no distinction between plural marking of direct and oblique arguments (examples from Mahayespur):

- /manus-ʈa/ 'man-CLF' = 'the man';
- /manus-ɛr baɽi/ 'man-GEN home' = 'the man's home'
- /manus-la/ 'man-PL' = 'the men';
- /manus-la-r baɽi/ 'man-PL-GEN home' = 'the men's home'

Chatterji (1926:731) finds this PL.OBL affix in Bangla to be 'well established by the end of the 15th century' on the basis of textual evidence. This feature of Bangla is innovative and unique—with no cognate affix found 'in any other NIA. language' (ibid.:730). It is also linguistically complex in its morphological conditioning and thus diagnostic of a propagation event including SCB, but excluding KRDS.

[MI 1.] > /-ḍ̲-/ 'PL.OBL.AN' {SCB} (before AD 1500). Diagnostic.[5]

---

[5] In formalising this morphological change, and those to follow, the following convention has been used: an innovation starting with '>' should be read as 'the following morpheme was introduced to the morphological system with function as given'. Thus [MI 1.] is to be read as 'the morpheme /-ḍ̲-/ was introduced with function PL.OBL.AN into SCB'.

The eastern Bangla varieties of Dhaka and Maimensingh use /-go/ with the same function as SCB /-ḍ-/ (P. Dasgupta 2003:365). This formally distinct innovation is also likely to be diagnostic of a PE.

The Oriya morpheme /-[ɔ]ŋ/ which marks oblique plural arguments is a retention of part of the MIA declensional morphology (see Chatterji 1926:723–724).

5.3.3 Nominative marking in eMg

The nominative case is a core grammatical function encompassing the S of intransitive clauses and the A of transitive clauses. In Bangla, NPs in nominative case are suffixed by /-[e]ra/ 'PL.NOM.AN' when the referrent of the head noun is both plural and animate. Oriya has a morpheme with nearly the same function: /-manɛ/ 'PL.NOM'. Note that the animacy criterion does not apply in Oriya. The function of these markers is independent of the transitivity of the verbal construction. Both Oriya and Bangla nominative plural markers are innovative and unique and thus diagnostic of (distinct) propagation events, based on linguistic complexity.

The Bangla affix /-[e]ra/ is etymologically linked with the genitive case, which is /-[e]r/ in Bangla and KRDS (see §5.3.5), and /-ər/ in western Magadhan lects. Chatterji (1926:734) writes:

> Originally, there was a noun of multitude after the strengthened genitive in « -ā ». This stage is still found in the Maithili « hamarā-sabh … » we …; and in Bengali, the noun of multitude can be optionally used

The 'strong' form of the genitive (that is, suffixed with /-a/, cf. §4.4.11), followed by a 'noun of multitude' (such as 'all') is found as a marker of plurality in early Maithili (Jha 1985 [1958]:389) and early Asamiya (Kakati 1962:294). The 'weak' form of the genitive (without the suffixed /-a/) is found with the same function in Magahi /-ər-ni/, and Bhojpuri /-ər-ən/ (Chatterji ibid.:734–736).[6] In sum, the marking of nominal plurality through a construction *noun-GEN(*-a*) + 'noun of multitude'* is well distributed in Magadhan lects and seems to been inherited from the proto Magadhan stage of linguistic history.

A morphological innovation occurred when the noun of multitude was *left off* from this plural construction *only in the personal pronouns*, without a change in function.

[MI 2.] *pronoun-GEN(-a) + noun of multitude* 'plural pronoun' > *pronoun-GEN(-a)* 'plural pronoun' {middle Bangla, early Asamiya, KRDS}. Supportive, not diagnostic.

That is, the inherited construction became *pronoun-GEN*-a, but retained plural function (even without the noun of multitude). It is important that this genitive-related suffix was initially grammaticalised with plural function *only as part of the pronominal system.* The middle Bangla literature provides evidence that [MI 2.] had occurred as early as the 14th century (ibid.). KRDS maintains the morpheme /-ra/ as the plural marker in nominative pronouns.

Modern Asamiya constructs its plural pronouns somewhat differently, but there is textual evidence that in early Asamiya the situation was the same as pertains in KRDS and Bangla today. The early Asamiya plural pronouns were suffixed with *-rā* followed by a numeral (rather than a noun), for example *torā dukānta* 'you:PL two' = 'the two of you'

---

[6] The extension of nominal stems with *-a has been discussed in §4.4.11; it is not unique to eMg and its diagnostic value for propagation events has not been reconstructed in this study.

(Kakati 1962:294). Recall from §5.2 that Asamiya, Bangla, and at least some KRDS lects all allow this post head position for numerals to indicate definite pragmatic function. From the existence of the construction *noun-GEN*-a + *NUM* it is not difficult to see how the numeral may be left off, leaving only *noun-GEN*-a as a construction with plural function.

Goswami makes the case (in his editorial comments to Kakati 1962:294, fn.2) that the function of this morpheme /-ra/ had already shifted from GEN to plural *by the time of early Asamiya*. He gives two lines of argument. Firstly, the noun phrase could be suffixed with the standard genitive marker *in addition to* /-ra/, as in the following example: *tārā dui-**r** ān nāi* 'they-*ra* two-GEN other not' = 'both of them have none else'. In this clause the genitive case *-r* marks the function of the NP whose head is not *dui* but *tā-* 'they'. One head noun can only take one case marker, therefore *-rā* can be considered to have ceased to function as a case marker in early Asamiya. Secondly, there are instances in early Asamiya writings where the morpheme *-rā* is suffixed with dative case: for example *tārāk* 'him'. The *-rā* element had clearly lost its erstwhile genitive function, because the head noun is in dative case (indicated by *-k*) not genitive case.

The shift in function of /*-ra/ from genitive to plural *in the personal pronouns* (formalised by [MI 2.]), involves a reduction in the complexity of the construction—which counts against its diagnostic value for a propagation event (cf. §3.4.1.1). Nonetheless, the conditioning of the change involves a change in function restricted to pronouns, which is thus relatively complex. The diagnostic value is registered as presently unclear, based on the criterion of linguistic complexity. It may be supportive of a propagation event diagnosed on other grounds.

Middle Bangla documents of the 15th century show that in Bangla at least, the scope of this innovative affix /-ra/ 'NOM.PL' had by that time expanded beyond personal pronouns to nouns in general. This change is peculiar to Bangla and does not characterise either early Asamiya or present-day KRDS:

[MI 3.] /-[e]ra/ 'PL.NOM' in pronouns > /-[e]ra/ 'PL.NOM.AN' in general nominal morphology {SCB} (by the C15th). Diagnostic.

The reinterpretation of this morpheme /-[e]ra/ as a marker of plural subjects (i.e. no longer restricted to the pronouns) is unique to Bangla in the Mg lects and diagnostic of a PE. This Bangla change is the nominative counterpart of [MI 1.] which introduced marking of oblique plural nouns. Together these two changes constitute a partial restructuring of nominal declensions in Bangla which is not shared with KRDS or Asamiya.

5.3.4 Ergative marking in eMg

Ergativity is a complex matter in NIA. In KRDS, as in Oriya and Bangla, overt marking of the Agent in transitive clauses has been completely lost, though it is attested in earlier stages of Oriya and Bangla. The Agent of transitive clauses is marked by a suffix in Asamiya, but it is usually termed the 'agentive' or 'nominative' case because the suffix does not affect agreement marking on the verb (which always agrees with the subject regardless of transitivity). The ergative-absolutive construction was replaced with a nominative-accusative construction concurrently with the addition of subject agreement endings on past and future tense formations (cf. §6.4). The loss of ergativity is a common feature of many NIA lects and its value for diagnosing unified propagation events is uncertain given its far reaching range over NIA. The exception to this loss is the maintenance of ergative/agentive marking in Asamiya—which is plausibly connected to

contact with speakers of ergative marking Tibeto-Burman languages (cf. Masica 1991:339ff.). In both cases (that is, the loss of ergativity, and its maintenance) the changes are nondistinctive in their respective linguistic ecologies, and non-diagnostic of PEs.

5.3.5 The genitive case marker

The genitive case is cognate in all eight KRDS lects but with some phonological differences between the sites (see Table 5.1). This charted representation of Magadhan data will be used throughout the chapter. In the chart the eight KRDS lects are separated from other lects by a single line. Superscript '$^{V}\_$' should be read as 'after a vowel', and '$^{C}\_$' as 'after a consonant'.

Table 5.1: Genitive forms in KRDS and some other NIA lects

Tharu -ək	RL $^{V}\_$ -r, $^{C}\_$ -ɛr	MH $^{V}\_$ -r, $^{C}\_$ -ɛr	SH $^{V}\_$ -r, $^{C}\_$ -ɛr
°SCH [a] -ka, -ki, -ke	KS $^{V}\_$ -r, $^{C}\_$ -ɛr	TH $^{V}\_$ -r, $^{C}\_$ -er	BH $^{V}\_$ -r, $^{C}\_$ -er
°Awadhi -ker, -ki, -kae	Kharia Thar $^{V}\_$ -r, $^{C}\_$ -ɔr	RP $^{V}\_$ -r, $^{C}\_$ -er	BN $^{V}\_$ -r $^{C}\_$ -ɛr; ɔr
Maithili -ker (-k)	Mal Paharia $^{V}\_$ -r, $^{C}\_$ -er	Rajshahi $^{V}\_$ -r, $^{C}\_$ -er	Kamrupi $^{V}\_$ -r, $^{C}\_$ -ɔr
Bhojpuri -kæ	Lodha $^{V}\_$ -r, $^{C}\_$ -er	SCB $^{V}\_$ -r, $^{C}\_$ -er	SCA $^{V}\_$ -r, $^{C}\_$ -ɒr
	Manbhum Paharia $^{V-}$ -r, $^{C}\_$ -er	Oriya -rɔ	

[a] The symbol ° indicates, following Masica, a preceding oblique linkage—that is an oblique marker which links the case marker and the noun.

The only morphological divergence within KRDS is the grammatically unconditioned variation in BN between two forms: /-ɛr/ and /-ɔr/. The -[ɛ]r variant is shared with other KRDS lects, the -[ɔ]r variant is shared with SCA and Kamrupi Asamiya, as well as Kharia Thar to the south-west of KRDS. Both variants, /-ɛr/ and /-ɔr/, have their origins in, and are inherited from, the common Magadhan stage of linguistic history.

During late MIA, the inherited OIA genitive affixes were lost through phonological reduction. However, before their complete loss, certain postpositions came into use as 'help words' for establishing the genitive function of NPs. These postpositions are the source of genitive case markers in most of NIA today, and KRDS is no exception. The two postpositions attested in MIA which pertain to this discussion are *kēra* and *kara*. The former is reconstructed by Chatterji as a semi-tatsama form of OIA *kārya*, with transposition and reduction during MIA > **kāira* > *kēra*. Derivatives of this etymon are found throughout Magadhan, as well as further afield, notably 'in the speech of European Gipsies who went with their language from North-Western India during the Second MIA. period' (Chatterji 1926:753). Later, (ibid.:755) he writes:

> Side by side with «kārya > kēra, kēla», the words «kara, kāra» … were used in Māgadhī Prakrit and Apabhraṁśa to indicate the genitive. It would seem that in Māgadhī Ap. «kara» was used with the pronoun originally, and then was extended to the noun … [In Mg lects] «kēra, kara» have become practically doublets of an identical genitive affix.

These two forms have existed as variants with the same grammatical function since late MIA. The selection of either one of these two historical variants is a case of *inheritance of variation with subsequent regularisation* (cf. §3.4.1.4). This type of change is not diagnostic of a propagation event because of the possibility of independent regularisation of the variation. The presence of /-ɔr/ in Asamiya, but /-ɛr/ in Bangla and KRDS, suggests that this variation was still present during the common Asamiya-Kamta stage (proto Kamrupa, pre-13th century). This hypothesis is confirmed by early Asamiya writings which for genitive case use: *-kera*; *-era*; *-kara*; *-ka* (Kakati 1962:306). (Note that single medial *k* was undergoing deletion in this environment during late MIA and early NIA).

There are two different explanations for the general presence of the -[ɛ]r genitive across KRDS but variation between -[ɛ]r and -[ɔ]r in BN. Firstly, proto Kamta may have inherited the -[ɛ; ɔ]r variation from proto Magadhan and proto Kamrupa. BN then, alone of the eight KRDS lects, may have retained the variation through to the present day. Secondly, the -[ɛ]r variant may have been regularised as the unique Genitive marker during the proto Kamta stage. The presence of -[ɛ; ɔ]r variation in present-day BN could be the result of a mixed inheritance—incorporating material from both proto Asamiya *-[ɔ]rɔ and proto Kamta *-[ɛ]rɔ.

This mixing of linguistic ancestries in BN may be recent or may have existed for centuries. We cannot exclude the possibility that this variation has continued unbroken in BN's linguistic ancestry since the proto Magadhan stage. Nonetheless the absolute absence of the -[ɔ]r variant in KRDS outside of BN suggests it is more likely that this variant was re-introduced into BN through its phylogenetic reintegration with Asamiya (cf. §7.5.4.2). The proto Kamta form is therefore reconstructed as *-[ɛ]rɔ.

[MI 4.] > /-[ɔ]r/ 'GEN' {BN, from SCA}. Supportive of contact relations with Asamiya.

Determining the sequencing of this reintroduction of variation is more difficult. We shall delay judgement until all innovations have been reconstructed, then use the less ambiguous aspects of BN's linguistic history to interpret the more ambiguous areas such as this.

Regardless of whether the reintroduction occurred at an early or recent stage of BN's linguistic history, the fact remains that BN incorporates several features—both morphological and phonological—that have been introduced from Asamiya. Alongside what might be called BN's proto eastern Kamrupa ancestry (i.e. Asamiya ancestry), there are some morphological features—notably in pronominals—which distinguish BN from Asamiya and instead associate it with proto Kamta. For this reason, BN is considered a transitional lect—intermediary between Asamiya and the other KRDS varieties.

Even in a dialect continuum, not all lects are equally transitional. There are centres of innovation and stability, which contrast with more variable areas that mix and match the features of adjacent lects. BN is a transitional lect; its linguistic ancestry is mixed, with some innovative features traceable to the proto Kamta stage, and others traceable to the proto eastern Kamrupa (Asamiya) stage.

For the reasons given above, proto Kamta is reconstructed as having regularised the *-ɛrɔ variant at the expense of the *-ɔrɔ variant. Note that this regularisation of variation, while reconstructed as part of the proto Kamta stage, is not *diagnostic* of that stage. It is a

principle of this reconstruction that the regularisation of inherited variation is not diagnostic of a propagation event because of the possibility that regularisation could have taken place independently with the same structural-linguistic outcome.

[MI 5.] Regularisation of *-[ɛ]rɔ in genitive function {KRDS, …}. Nondiagnostic.

Table 5.2: Genitive case forms reconstructed for various stages of NIA history

	Reconstructed forms
p-eMg	*-kara*; *-kēra*
pKmt	*-ɛrɔ [a]
p-eKamrupa (Asamiya)	*-ɔrɔ

[a] Becomes /ɛr/ by loss of final *ɔ, cf. §4.4.11.

There is therefore no diagnostic value to the cognacy between the proto Kamta form and the Bangla form /-er/, or for that matter any eIA genitives which are reflexes of the *kera* variant instead of the *kara* variant. The variation was inherited, and regularisation was plausibly independent (cf. §3.4.1.4).

5.3.6 Dative-accusative marking

The dative-accusative (DAT) case has three major uses in KRDS, as in many other NIA lects: first, to mark the recipient of a ditransitive verb, for example 'I gave the book *to the boy*'; second to mark the object of a transitive verb ('P' in Comrie 1978) whose referent is either [human] or [animate and discourse prominent], for example 'I saw *the boy*';[7] third to mark logical subjects which have the semantic role of experiencer as in the example below (from Mahayespur).

/mo-**k** duk^{h} lag-i-c-e/
1SG-**DAT** fear attach-PFV-PRS-3
'I feel scared' or: 'Fear has struck me!'

A proper syntactic description of the grammatical relations involved in this construction is beyond the scope of this study.[8] Note that in Bangla the experiencer in this construction is marked with *genitive* not dative case.

Using the same case to mark both objects of transitive verbs and indirect objects of ditransitive verbs is a common feature of NIA. Masica (1991:365) observes:

> Historically, the Indo-Aryan Accusative merged with the Nominative … This is not to say that Objects are always bereft of case marking. They may take it, in the form of the Dative marker.

The only lect in Table 5.3 which has distinct forms for accusative and dative functions is Manbhum Paharia (according to D. Dasgupta 1978:248). In all other lects the one case marks NPs with both dative and accusative function.

---

[7] Wilde (2008:110) points out that even animate and human direct objects can be unmarked for Dative-Accusative if they are *specific indefinite.*

[8] For a general description of the dative subject construction from a pan-NIA perspective see Masica (1991:346ff.), for papers giving details for individual NIA lects see Verma and Mohanan (1990). See Wilde (2008:113–115, 305–306) for a morphosyntactic description of the dative subject construction in central Jhapa Rajbanshi (neighbouring to MH, used in this study).

The dative-accusative case marker is cognate at all eight KRDS sites; it is the phonologically regular reflex of proto Kamta *-ɔkɔ.

Table 5.3: Dative-Accusative forms in KRDS and some other NIA lects

Tharu -hənə	RL $^{V\_}$-k, $^{C\_}$-ɔk	MH $^{V\_}$-k, $^{C\_}$-ɔk	SH $^{V\_}$-k, $^{C\_}$-ɔk
SCH -ko	KS $^{V\_}$-k, $^{C\_}$-ɔk	TH $^{V\_}$-k, $^{C\_}$-ok	BH $^{V\_}$-k, $^{C\_}$-ok
Awadhi [-kə, -ka] [a]	Kharia Thar $^{V\_}$-k, $^{C\_}$-ɔk	RP $^{V\_}$-k, $^{C\_}$-ok	BN $^{V\_}$-k, $^{C\_}$-ɔk
Maithili -kẽ˜	Mal Paharia $^{V\_}$-k, $^{C\_}$-ek	Rajshahi $^{V\_}$-k, $^{C\_}$-ok	Kamrupi $^{V\_}$-k, $^{C\_}$-ɔk
Bhojpuri -ke	Lodha -ke	SCB -ke, -ḍer [b]	SCA $^{V\_}$-k, $^{C\_}$-ɒk
Konkani -k(ə), $^{PRO\_}$-ka	Manbhum Paharia ACC: -ke DAT: $^{V\_}$-k, $^{C\_}$-ek	Oriya -ku	

[a] These variants are phonologically conditioned (Masica 1991:245).

[b] /-ḍer/ is a general oblique ending for plural animate nouns, see §5.3.2.

The prosodic raising of *ɔ > /o/ affects the case marker in TH, BH and RP (cf. §4.4.4). The chronology of this change is argued in §7.5.3.2 to be post AD 1800 on the basis of sociohistorical sequencing. This chronology places the raising innovation subsequent to the proto Kamta stage, and thus the proto Kamta form in Table 5.4 is reconstructed with the lower vowel: *-**ɔ**kɔ.

Dative-accusative forms along the lines of -(V)k(V) may be found in most NIA languages. Several of these are likely cognates of KRDS *-ɔkɔ, though Masica writes 'It is not ... clear whether [all NIA datives based on /k/] are to be ascribed a common origin' (1991:245). The etymology is ambiguous, with Chatterji citing three or four different possibilities.

SCA and Kamrupi forms are clearly cognate with KRDS *-ɔkɔ. Chatterji (1926) reconstructs the SCB /-ke/ as the result of agglutination of *-k 'DAT' + *-ɛ 'INS-LOC'. Given that the closely related lects SCA and KRDS have regular reflexes of *-ɔkɔ, the SCB etymology should be slightly amplified as follows: -ke 'DAT' < *-[ɔ]ke < *-ɔkɔ 'DAT' + *-ɛ 'INS-LOC'.[9] The extension of the dative with the instrumental-locative suffix /-e/ mirrors the extension of the Bangla locative to a double locative (see §5.3.7). It is found beyond Bangla in other eastern Magadhan lects, for example Lodha. Before this innovation can be considered diagnostic of a propagation event, its broad dialectology in eMg needs to be studied and showed to be conducive to sociohistorical explanation.

[9] This fuller etymology is implied but inexplicit in Chatterji's analysis because his description gives Romanised transliterations of the written form, rather than phonological forms, and /-ke/ and /-[ɔ]ke/ are homographs in Bangla script.

[MI 6.] *-ɔkɔ 'DAT' + *-ɛ 'INS-LOC' > /-ke/ 'DAT' {Bangla, …}. Supportive, not diagnostic.

The chronology of [MI 6.] in Bangla is not clearly stated by Chatterji. He writes that the older form /-ɔk/ is 'exceedingly common in M.B., as in the ŚKK. and other works' (ibid.:759), but then that '«-kē» occurs regularly in MB., NB.;' (ibid. 762).

This discussion suggests the reconstruction of the following forms for some post-Magadhan stages of linguistic history pertinent to KRDS's history.

Table 5.4: Dative-Accusative case reconstructed for various stages of NIA history

	Reconstructed forms
p-eMg	?
pKamta	*-ɔkɔ [a]
p-eKamrupa (Asamiya)	*-ɔkɔ
pGauḍa-Baŋga (Bangla)	*-ɔkɔ + *-ɛ > /-ke/

[a] Becomes /ɔk/ by loss of final *ɔ; cf. §4.4.11.

As the prior ancestry of *-ɔk(ɔ) is ambiguous, it is unclear whether Asamiya-Bangla-KRDS *-ɔkɔ constitutes an innovation diagnostic of a pre-Gauḍa-Kamrupa propagation event or not. The Oriya form is /-ku/ which is probably (but not conclusively) cognate. The KRDS and Asamiya forms while identical are inherited morphemes, partially cognate with the -k element of (at least) Bangla /-ke/ 'DAT'. As the Asamiya-Kamta similarity in this feature is not innovative, it cannot be used as diagnostic for the common proto Kamrupa (Asamiya-Kamta) stage.

Before moving to locative marking, we may briefly visit the evidence for earlier dative-accusative marking as found in the Buddhist mystical songs, the *Caryāpadas*. Chatterji writes that the *-ɔk form is 'used for the dative in [Old Bangla]', that is in the *Caryās* (ibid.:759), but then later adds that 'The Caryās, in addition to «-ka» and «-ku», give instances of «-kē».' (ibid.:762) This may suggest a history of inherited variation with subsequent regularisation, which would entail that SCB /-ke/ ([MI 6.]) is also not diagnostic of a propagation event. However, it must be kept in mind that the sentence structure of the Caryā songs is intentionally poetic and that extended morphology in the songs may reflect considerations of rhyme or meter rather than a single coherent vernacular of the time. Anyhow, the status of the evidence in the Caryās is so complex and controversial—'deliberately enigmatic', as Masica (1991) puts it—that P. Dasgupta's (2003) policy of agnosticism towards the classification of this 'language' (if, indeed, the songs reflect anything like a synchronically unitary lect) seems the most advisable course of action at the present time.

5.3.7 Locative marking

The locative marker, used to indicate locational adjuncts, is cognate across the eight KRDS sites. The forms differ only due to prosodic vowel raising.

Table 5.5: Locative case markers in KRDS and some other NIA lects

Tharu -mə	RL V_ -ṭ, C_ -ɔṭ	MH V_ -ṭ, C_ -ɔṭ	SH V_ -ṭ, C_ -ɔṭ
SCH meˇ, pər	KS V_ -ṭ, C_ -ɔṭ	TH V_ -ṭ, C_ -oṭ	BH V_ -ṭ, C_ -oṭ
	Kharia Thar -e V_ -k, C_ -ɔk	RP V_ -ṭ, C_ -oṭ	BN V_ -t, C_ -ɔt
Maithili -e, me, -hi, -tə	Mal Paharia V_ -ṭ, C_ -eṭ		Kamrupi -ɔt
	Lodha -e, -ke, -ṭe	SCB C_ -e, V_ -ṭe [a]	SCA -ɒt
Marathi -ī, -ṭ	Manbhum Paharia V_ -e, C_ -ɔe V_ -k, C_ -ek	Oriya -re	

[a] In SCB the allomorph /-e/ is used after consonants, and optionally after non-high vowels. The historically 'double locative' allomorph /-ṭe/ is mandatory after a high-vowel and optional after non-high vowels (P. Dasgupta 2003:364).

Analogously to the dative-accusative above, the KRDS locative is cognate with the SCA form /-ɒt/. SCB has two allomorphs with locative function: the allomorph /-ṭe/ is partially cognate with SCA and KRDS, while the allomorph /-e/ is not cognate. Early Oriya *-e* is cognate with the latter Bangla allomorph (Chatterji 1926:746). The modern Oriya locative marker is /-rɛ/, which is not cognate with KRDS /-ɔṭ/.

The ancestry of this *-t-* based locative is an open question. Chatterji (ibid.:750) writes:

> The «-ta» postposition characterises the Bengali-Assamese group only among Magadhan speeches. This postposition would nevertheless seem to have been a Magadhan (MIA.) inheritance in Bengali. At the present day, it occurs as «-ta» in Assamese and in dialectal Bengali (North, East, South-east).

The absence of a locative in /ṭ/ in modern Oriya or earlier Oriya documents casts some doubt over the presence of *-ɔṭɔ 'LOC' during the common proto-eastern Magadhan stage (ancestral to Oriya, Bangla, Asamiya and KRDS). However, when we look further afield, a potentially cognate *-t-* based locative is found in the more distantly related lects, Maithili and Marathi. According to Masica, the *-t-* based locative is 'derived from the older locative postposition *-ta* (cognate with M. [Marathi] *-ā(˜)t* < OIA *antar* 'the inside')' (1991:213).[10] According to Jha (1985 [1958]:34–35) a cognate morpheme is found in Maithili linguistic history:

> The presence of the loc. forms in *-ta* in the *Caryās* suggests, at first, a connection with Bengali. But *-ta* in extended forms is met with in early literary Maithilī as well as in the modern eastern Maithilī dialect: of course, even there, it is no longer comonly

[10] Italicised forms are transliterated orthographic representations. Kakati (1962:305) disputes the OIA etymology proposed by Masica.

> used. … As a matter of fact, it may be considered to be a loc. affix current in the whole of northern, central, and eastern Māgadhan area inasmuch as it occurs in Assamese, Maithilī and Bengali

Given that cognates of this locative morpheme are distributed beyond Bangla-Asamiya-KRDS and found with the same function in Maithili and possibly also in Marathi, the heritage of the -t- based locative seems to stretch back in time beyond the hypothetical proto Gauḍa-Kamrupa stage, to the hypothetical proto Magadhan stage. The absence of a cognate morpheme in Oriya must then be explained by proposing the *replacement* of this inherited locative at quite an early stage of Oriya's linguistic history. This hypothesis should be considered tentative until subjected to testing based on a broader sampling of data from the Magadhan lects.

Table 5.6: Locative case forms reconstructed for various stages of NIA history

	Reconstructed case forms
pMg	?*-ɔt̪ɔ
p-eMg	?*-ɔt̪ɔ, written *-ত, Romanised as *-*ata*
pKmt	*-ɔt̪ɔ
p-eKamrupa (Asamiya)	*-ɔtɔ
pGauḍa-Baŋga (Bangla)	*-ɔt̪-ɛ *-ɛ

At any rate, the double locative found in Bangla is certainly innovative. There is no clear evidence that both *-ɔt̪ and *-ɔt̪-ɛ occured during the proto-eastern Magadhan stage or earlier for that matter (unlike for the variation in genitive case *-kera, -kara*).[11] The testimony of the Sri Krishna Kirtana manuscript is that this locative doubling innovation occurred early in Bangla linguistic history—either during or prior to the 14th Century. [MI 7.] is likely to be diagnostic of a propagation event:

[MI 7.] *-[ɔ]t̪ 'LOC' + *-ɛ 'LOC-INS' > /-t̪e/ 'LOC' {SCB, Lodha} (before AD 1400). Probably diagnostic.

To summarise the key points: locative markers in KRDS and SCA are cognate, and partially cognate with SCB /-t̪e/. The doubling of the SCB locative is probably diagnostic of an PE, but the Asamiya-KRDS locatives are inherited and the structural similarity is not diagnostic of a change event.

5.3.8 Instrumental marking

At this point the analysis shifts from phonologically bound inflectional case suffixes, to postpositions which occur directly after the noun, without an intervening genitive marker. The categories of postposition, clitic and suffix are natural points along the pathway of

---

[11] Analagously to the history of the dative outlined above, Chatterji mentions 'one or two instances' in the *Caryās* of 'double locative' forms, but 'numerous instances' of the locative *-ta* [-ɔt̪ɔ] (ibid.:750). For the reasons sketched under §5.3.6, this reconstruction maintains an agnosticic position on how to interpret the data found in the *Caryās*, especially concerning whether or not the data constitute a unitary historical lect. Data from the Caryās alone are not sufficiently reliable to establish inherited variation with subsequent generalisation. Therefore the extension of the locative *ɔt̪ɔ with *ɛ stands as a PE-diagnostic change.

grammaticalisation, and it is not surprising that older suffixes are supplanted by newer postpositions, which in turn are phonologically reduced as they move towards suffix status. Instrumental forms for KRDS were collected using the sentence frame 'I am writing with a pen', and are compared with data from other NIA lects in Table 5.7.[12]

Table 5.7: Instrumental forms in KRDS and some other NIA lects

Tharu	RL	MH	SH
le, leka	sɛ	**ḍɛ**	**ḍi**
SCH	KS	TH	BH
se	sɛ	**ḍehene**	**ḍiæ**
	Kharia Thar	RP	BN
	-e, -**ḍie**	**ḍiæ**	$^{V\_}$*-re*, $^{C\_}$*-ere*
Maithili	Mal Paharia		Kamrupi
-e, e˜, sə˜, **ḍea**	-e, -herɔĩ, -hilẽ		-e-**di**
	Lodha	SCB	SCA
	-e, -**ḍi**	-ṭe [a], -**ke ḍie**	-e, -*er*-e [b], **di**, -e-**di** [c]
Marathi	Manbhum Paharia	Oriya	
-e˜, - ī ˜, -ne˜, n ī ˜	-e˜, -**ḍia**	-e, -*re*, **ḍei**	

[a] The instrumental-locative is used for inanimate instruments (Masica 1991).
[b] Masica (1991:246).
[c] Kakati (1962:304).

The KRDS data in Table 5.7 include variants of three etyma, which are referred to here as s-, d- and r- based instrumentals. The s-based instrumental is found in RL and KS, as well as in Hindi. The *s*-based instrumental is not found in Asamiya-Bangla-KRDS-Oriya (the eastern Magadhan lects) outside of RL and KS. This distribution of the s-based instrumental suggests it is a Hindi loan.

[MI 8.] > /sɛ/ 'INS' {RL, KS from Hindi/Bihari}. Diagnostic of contact relations of diglossia with Hindi, not a PE.

As diglossic contact with Hindi is a likely conditioning factor for this change, the range (joining RL and KS) is not diagnostic of a propagation event.

Turning to the d-based instrumental, this etymon is found repeatedly in eastern Magadhan, see Table 5.7 where d-based instrumental markers are in bold face. This etymon is the perfect participle form of the verb *ḍɛ- 'give'. Modifications to the suffix *-ia > /-ɛ,-i,-hene/ are due not to nominal morphological changes but to phonological changes and verbal morphological changes.

The range of this etymon (d-based instrumental) throughout eastern Magadhan, and also in Maithili (central Magadhan) suggests it is an inherited feature from pre-eastern Magadhan, and is accordingly listed in Table 5.8. This accords with Chatterji's judgement that this form is 'Found from early times: e.g. ŚKK' (ibid.:770). The presence of this etymon in post-Mg lects is an inheritance, and not the result of a change event.

[12] Some KRDS lects have an inflectional instrumental /-ɛ/ for inanimate instrumentals (similar to Asamiya and Maithili *-e*, and Bangla /-ṭe/) but the relevant data have not yet been systematically collected for all lects.

The r-based instrumental is found in eastern Magadhan lects Oriya and SCA among others (see the italicised forms in Table 5.7). This discontinuous range suggests that the r-based instrumental is inherited from at least the proto-eastern Magadhan stage. Misra connects this marker etymologically to the inherited genitive marker /-ɛr; -ɔr/ extended by the locative-instrumental ending /-e/ (1975:61).

With respect to the presence of the r-based instrumental in BN of KRDS, two explanations are possible (as in §5.3.5 above). The r-based instrumental may have be inherited into the proto Kamta stage and lost at all KRDS sites excepting BN. Alternatively, it may have been replaced by the d-based instrumental at the proto Kamta stage, and the presence of the r-based instrumental in BN resulted from its dual Kamta+Asamiya linguistic ancestry. The latter explanation is the more plausible and economical reconstruction. As argued in §5.3.5, some morphological features of BN are clearly inherited from proto Kamta, with others clearly from proto-eastern Kamrupa (Asamiya); others yet are ambiguous between the two. The r-based instrumental is one of those features inherited into BN not from proto Kamta, but from proto-eastern Kamrupa (Asamiya).

The proto Kamta stage is reconstructed without the *-[ɛ; ɔ]r-ɛ instrumental marker. While this loss is held to have occurred during the proto Kamta stage, it is not a diagnostic feature of this stage. As for other changes involving the loss of inherited variants, the loss of *-[ɛ; ɔ]r-ɛ 'INS' is not a good diagnostic for a propagation event (cf. §4.4.1.1).

The instrumental case forms reconstructed for various stages of eastern Magadhan linguistic history are summarised in Table 5.8.

Table 5.8: Instrumental case forms reconstructed for certain stages of NIA history

	Reconstructed case forms	
p-eMg	*d̯ia	*-[ɛ; ɔ]r-ɛ
pKmt	*d̯ia	
p-eKamrupa (Asamiya)	*d̯i	*-[ɛ]rɛ
pGauḍa-Baŋga (Bangla)	*d̯iɛ	

Both d-based and r-based instrumental markers are inherited from proto-eastern Magadhan, while the s-based instrumental occurs in KS and RL due to borrowing from Hindi. The presence or absence of d-based and r-based instrumental markers are not diagnostic of propagation events because they constitute inheritance of variation with subsequent regularisation. The *r*-based instrumental is not reconstructed for proto Kamta as its presence in BN is more simply explained by reference to BN's proto-eastern Kamrupa (Asamiya) ancestry.

5.3.9 Ablative and Comparative marking

The functions of Ablative and Comparative are grammatically interrelated in eastern NIA in general, including KRDS. For this reason, forms which serve either or both of the two functions are analysed concurrently in this section. Markers of both these functions are postpositions of sorts, occurring either after the noun head, or after the head suffixed by genitive case. However, there is a general asymmetry in the relation between markers for ablative and comparative. The ablative form may be used for comparative function, but in

most KRDS lects (not RL, KS, MH) there is a uniquely comparative postposition, which may not in turn be used for general ablative function.

Ablative forms for the KRDS sites are shown in Table 5.9. These were collected using the sentence frame 'Ram/Mohammed fell from the tree'. KRDS comparative forms were collected using a frame such as 'I am taller than you, he is taller than me, etc.'.

Ablative marking is highly fragmented in eastern Magadhan. Historical texts record an Apabhraṁśa (late MIA) ablative suffix [*huˇ*; *-hu*], which Chatterji reconstructs as also inherited into eastern (Magadhi) Apabhraṁśha. The only evidence of inheritance into eastern Magadhan lects is Oriya /-u/, and possibly Lodha /-nu/. In other eastern Magadhan descendants this suffix has been replaced by a range of new forms.

The geographically central KRDS sites are alike to Bangla lects in employing the verb root /ʈ̪ʰak-/ 'stay, remain' in perfect participial form. Chatterji finds this form as early as the ŚKK (before AD 1400), but no earlier. The restricted distribution of this form in Magadhan lects—not found beyond KRDS and Bangla—points to an innovation rather than an older inherited form.

[MI 9.] > *ʈ̪ʰakia 'ABL' {SCB, TH, SH, RP, BH}. Tentatively diagnostic of contact relations with SCB through diglossia.

Table 5.9: Ablative and Comparative forms in KRDS and some other NIA lects

Tharu se	RL ABL sɛ CMP -GEN sɛ	MH ABL sɛ CMP -GEN sɛ	SH ABL haʈ̪ɛ, ʈ̪ʰaki CMP -GEN tsaja
SCH se, -GEN karən -GEN ʈ̪ərəf se	KS ABL sɛ CMP -GEN sɛ	TH ABL ʈ̪ʰeke CMP -GEN ʧeje	BH ABL haʈ̪e, ʈ̪ʰaki CMP -GEN tsaja
	Kharia Thar ABL heˇʈ̪e ABL CMP -huˇ CMP -GEN lou, le CMP -GEN ʈ̪ʰaki	RP ABL ʈ̪ʰaki; ʈ̪ʰæki CMP -GEN sæja	BN ABL -GEN pɔra CMP -LOC kɔi
Maithili səˇ, -k karəne	Mal Paharia ABL -GEN ʧalaiˇ ABL -GEN niˇ		Kamrupi ABL -GEN pɛre, pai ABL -GEN pɛrai CMP -LOC ke, kori CMP -LOC ʈ̪ʰaki
	Lodha ABL -nu CMP -GEN ʧai	SCB ABL -ʈ̪ʰeke CMP -GEN ʧeje	SCA ABL -GEN pɒra CMP -GEN kɔi
Marathi -ūn, -hūn, -aʈ̪ūn, -mūleˇ, -pekṣ̪a	Manbhum Paharia CMP -GEN leˇ	Oriya -u, -ru, ʈ̪ʰaru, ʈ̪ʰiru	

This narrow distribution casts doubt over whether (i) the ablative marker *ʈ̪ʰakia should be reconstructed as part of the proto Kamta inheritance (with loss in KS, RL, MH and BN), or whether instead (ii) it was introduced into the other four lects (TH, SH, RP and BH) after the proto Kamta stage through borrowing from Bangla, or indeed whether (iii) it was borrowed into Bangla from these lects. This is a problem of sequencing which cannot be disambiguated on purely linguistic grounds. The range within KRDS of the ablative marker /ʈ̪ʰakia/ is limited to the Bengal sociopolitical zone, and for this reason [MI 9.] is labelled as 'tentatively' diagnostic of contact relations with Bangla. Some further discussion on this change comes in Chapter 7, but on the whole the history of this morpheme in KRDS remains unclear.

The absence of a stable ablative marker at the proto Kamrupa (Asamiya-Kamta) stage is supported by early Asamiya texts. These show nouns with ablative function marked as genitive followed by 'verbs implying removing, going away, descending' (Kakati 1962:309). Grammaticalisation of any of these verbs within the noun phrase had thus not occurred during proto Asamiya, nor during the still earlier proto Kamrupa stage.

The more western KRDS lects employ the same form for the ablative and comparative functions as for instrumental function: /sɛ/. This instrumental marker was diagnosed above as a borrowing from Hindi, and a similar explanation accounts for the ablative and comparative uses of this morpheme. In the case of [MI 10.], MH lect is also included in the range of the change event. For the same reasons as outlined in §5.3.8 for instrumental marking, this change is not diagnostic of a unified propagation event, but of diglossia with Hindi.

[MI 10.] > /sɛ/ 'ABL, CMP' {RL, KS, MH}. Diagnostic of contact relations through diglossia with Hindi.

The BN ablative and comparative forms are innovative for KRDS, and constitute borrowings from SCA. They are part of BN's proto-eastern Kamrupa (Asamiya) linguistic ancestry.

[MI 11.] > /pɔra/ 'ABL' {BN, SCA}. Diagnostic of contact relations with Asamiya.

[MI 12.] > /kɔi/ 'CMP' {BN, SCA}. Diagnostic of contact relations with Asamiya.

The form /haṯ̪e/ also occurs with ablative function in some KRDS lects, as well as in early Asamiya texts (cf. Kakati 1962.). It is a reflex of the OIA present participle of the verb 'to be': *santa*. Cognates are also found in Kharia Thar /-hɔ̃ṯ̪e/, Magahi /səṯ̪i/, Bhojpuri /sənṯ̪e/, and Middle Bangla *-hante; -honte; -haˇte; -hane.* This etymon did not constitute the ablative *suffix* for proto Magadhan (cf. rather Oriya /-u/); nonetheless it has been inherited from proto Magadhan as a postposition with some ablatival function. As an inherited feature, the /haṯ̪e/ postposition is not diagnostic of a morphological change event.

The presence of a postposition with uniquely *comparative* (and not simultaneously ablative) function is a feature of SCB and some of the KRDS lects. In these lects the comparative is based on the perfect participial form of the verb *ʧah- 'look at'. Unlike the ablative postposition *ʈ̪ʰak-ia, this comparative postposition is an inherited form for this function. Chatterji (1926:769) reconstructs the etymology as follows:

> চাহিযা, চেয়ে «cāhiyā > cēyē» *having looked at*, indeclinable conjunctive … used in comparison, with the genitive. This use seems to be old. Cf. Early Eastern Hindi as in Tulasī-dāsa.

The writings of Tulsidas show that this form was used with comparative function in a western Magadhan lect of the 16th century. This distribution—in both western Magadhan and eastern Magadhan—suggests it was inherited with this function from the common proto Magadhan stage, though the possibility of a more recent propagation cannot be completely ruled out. If cognate postpositions are found in more Magadhan lects, then this would strengthen the hypothesis that it is a postposition inherited with comparative function. Further reconstruction relating to this morpheme should investigate the syntactic processes that created this construction with comparative function, including the syntactic motivation for putting verbal participles after a genitive case-marked noun.

The reconstruction of ablative and comparative postpositions is summarised in Table 5.10.

Table 5.10: Ablative and comparative postpositions reconstructed for some stages of IA history

	Reconstructed forms	
	ABL	CMP
p-eMg	*-[huˇ; hu], *hante*	*ʧa-ja
pKmt	*haṯɛ ?	*ʧa-ja
p-eKamrupa (Asamiya)	*haṯɛ *pɔra	*kɔi
p-Gauḍa-Baŋga (Bangla)	*haṯɛ *ṯʰak-ia	*ʧa-ja

The sequencing of the propagation of *ṯʰak-ia ‘ABL’—whether before, during or after the proto Kamta stage—has been tentatively reconstructed to be *post-Kamta*, resulting from more recent Bangla influence in extended central KRDS (cf. §7.5.3.2). The other ambiguity registered in the table concerns whether or not *-[huˇ; hu] still pertained as ablative marker during the proto Kamrupa and then proto Kamta stages.

5.3.10 The reconstructed case system of proto Kamta, and its modern reflexes

The foregoing reconstruction of the proto Kamta system of case inflections is summarised in Table 5.11. Contemporary forms for the eight sample KRDS lects are given as reflexes in accordance with the discussion above. Putative borrowings (post-Kamta replacements) are shown by shaded cells.

Table 5.11: Reconstructed proto Kamta case system and its reflexes

	DAT	GEN	LOC	ABL
pre-pKmt	*-[ɔ]kɔ	*-[ɛ]rɔ	*-[ɔ]t̪ɔ	*hat̪e ?
RL	-[ɔ]k	-[ɛ]r	-[ɔ]t̪	sɛ
KS	-[ɔ]k	-[ɛ]r	-[ɔ]t̪	sɛ
MH	-[ɔ]k	-[ɛ]r	-[ɔ]t̪	sɛ
TH	-[o]k	-[e]r	-[o]t̪	t̪^{h}ɛke
SH	-[ɔ]k	-[ɛ]r	-[ɔ]t̪	hat̪ɛ t̪^{h}aki
RP	-[o]k	-[e]r	-[o]t̪	hat̪e t̪^{h}æki
BH	-[o]k	-[ɛ]r	-[o]t̪	t̪^{h}aki
BN	-[o]k	-[ɛ]r; -[ɔ]r	-[o]t̪	-GEN pɔra

Six morphologically-conditioned innovations have been reconstructed in this section, which together derive the contemporary systems from the reconstructed forms.

[MI 4.] > /-[ɔ]r/ 'GEN' {BN, from SCA}. Supportive of contact relations with Asamiya.

[MI 8.] > /sɛ/ 'INS' {RL, KS from Hindi/Bihari}. Diagnostic of contact relations of diglossia with Hindi.

[MI 9.] > *t̪^{h}akia 'ABL' {SCB, TH, SH, RP, BH}. Tentatively diagnostic of contact relations with SCB through diglossia.

[MI 10.] > /sɛ/ 'ABL, CMP' {RL, KS, MH}. Diagnostic of contact relations through diglossia with Hindi.

[MI 11.] > /pɔra/ 'ABL' {BN, SCA}. Diagnostic of contact relations with Asamiya.

[MI 12.] > /kɔi/ 'CMP' {BN, SCA}. Diagnostic of contact relations with Asamiya.

All of the innovations that affect KRDS case systems are replacements through (a) the influence of Hindi in the western KRDS lects {RL, KS, and to a lesser extent MH}, (b) the influence of SCA in the eastern KRDS lect {BN}, and possibly (c) the influence of SCB in the central lects {TH, SH, RP, BH} (though this last hypothesis is less robust than the others).

5.4 Specificity classification markers: description and reconstruction

5.4.1 Synchronic overview

Within the eastern Magadhan lects, there is a set of suffixes which attach directly to nouns, and specify the discourse status of the noun as either specific-indefinite or specific-definite depending on the relative position of the noun, numeral and classifier (cf. overview in §5.2, and Wilde [2008:63–84] for an indepth description of classification in central Jhapa Rajbanshi).

The inflectional categories marked by these suffixes are number and noun class, but noun class is only marked when the noun referent is singular and thus the plural marker is the same across all noun classes. KRDS differs in this regard from Bangla, which has

distinct plural markers depending on whether the referent is animate or not (this divergence established by [MI 3.]). The suffixes employed in Mahayespur (eastern Jhapa, Nepal) are given in Table 5.12 as an illustration of how all this works as a synchronic system (but see further Wilde ibid. for central Jhapa dialect).

Table 5.12: System of classifiers in Mahayespur (MH) of KRDS

		Noun class		
		Class I	Class II	Humans
Number	Singular	-ʈa	-k^{h}an	-ʤʌn [13]
	Plural	-la	-la	-la

Recall from §5.2 that classifiers may differ in their syntactic position within the NP. Classifiers /-ʈa/ and /-k^{h}an/ may occur in one of three positions in the noun phrase in MH: postnumeral, postdeterminer or postnominal (in that order of priority). The human classifer /-ʤʌn/ is only permitted in the postnumeral position in MH. (Wilde also finds it to occur after the indefinite pronoun [2008:71].)

The plural marker may occur postdeterminer or postnominal but not post-numeral, and plural marking is not permitted when the noun phrase includes a numeral. For example: /lok-la/ 'man-PL' = 'the men', or /tin-ʈa lok/ 'three-CLF man' = 'the three men'. The pragmatic implications of the position given to the classifier are considerably more complex than this, but this overview suffices for the purpose here.[14]

Some noun classes are mutually exclusive, while others may be subclasses of more general classes. In MH, all human referents are Class I nouns, thus three men can be either /ʈin-ʈa lok/ or /ʈin-ʤʌn lok/ 'three-CLF man' = 'the three men'. As mentioned below Table 5.12, the classifier /-ʤʌn/ is more restricted in its distribution than the other classifiers /-ʈa, -k^{h}an, -la/ and does not suffix to head nouns. Hence the definite singular of a human noun is, for example, /lok-ʈa/ 'the man', and not */lok-ʤʌn/. Based on this structural distinction between the single classifiers /-ʈa, -k^{h}an/ (which suffix to the noun) and /-ʤʌn/ (which does not), Wilde argues that only /-ʈa/ and /-k^{h}an/ should be considered nominal classifiers, and /-ʤʌn/ a numeral classifier (2008:71). Wilde thus considers there to be just two nominal classifiers in central Jhapa Rajbanshi.

The distribution of the /-ʈa/ and /-k^{h}an/ noun classes is partially motivated semantically, but for the most part semantically arbitrary (ibid.:70). The partial semantic motivation may be seen with the /-k^{h}an/ class, which coincides partially with the semantic criterion of flatness (Chatterji 1926:779). For example /kitap-k^{h}an/ 'the book', /duar-k^{h}an/ 'the door', etc. Wilde (2008:68) concurs that semantic features play 'at least some part' in assigning some nouns to the /-k^{h}an/ class:

- 'which classifier is associated with the noun पात pat 'leaf' seems to be determined by the size and shape of the leaf itself' (ibid.)
- 'the classifier -खान -*k^{h}an* is not used for animates … though it may be used for body parts' (ibid.)

---

13 In the Rajbanshi of central Jhapa, the corresponding morpheme is /-ʤ^{ɦ}ʌna/ (Wilde 2008:64).

14 For further synchronic details, see P. Dasgupta's (2003) analysis of the pragmatic effects of the syntactic position of classifiers in Bangla, and Wilde's (2008:78–84) description of central Jhapa Rajbanshi.

- 'The classifier -खान *-kʰan* tends to be used where there is an association with liquid or airborne substances.' (ibid.)
- However, 'both classifiers can be used for abstract entities.' (ibid.)

The /-ʈa/ class is rather more of a 'default' noun class, into which all left over nouns are thrown. In western Jhapa and Morang districts of Nepal, the default classifier /-ʈʌ/ (with allomorph /-ɖʌ/) even attaches to *proper nouns*. Such suffixing does not occur elsewhere in KRDS. Some dialects of Jhapa district allow double marking within the noun phrase, with the classifier attaching both to the head and to a numeral (Wilde 2008:76; note that Wilde cites this observation as requiring further confirmation through dialectal study). Double marking is ungrammatical in most KRDS lects.

Bangla has variants of the /-ʈa/ and /-kʰan/ suffixes which are conditioned by semantic and pragmatic factors: /-ʈa/ versus /-ʈi/ and /-kʰana/ versus /-kʰani/, the latter being basically a diminutive form (see further P. Dasgupta 2003:379ff.).

Table 5.13 lists the specificity-classifiers collected for this study at each of the eight KRDS sites, as well as cognate forms in other eastern Magadhan lects. These suffixes were collected using the nominal concepts glossed in Table 5.15. Cells are shaded if they contain forms that are not cognate with other forms in the same column.

Table 5.13: Specificity-classifers for singular noun referents in KRDS and some other eMg lects

	Specific-classification forms for singular noun referents								
RL	-[ʈʌ, ɖʌ]	-kʰan	-ʤʌn						
KS	-[ʈə, ɖə]	-kʰan	-ʤʌn						
MH	-[ʈə, ɖə]	-kʰan	-ʤʌn						
TH	-[ʈa, ɖa]	-kʰan	-ʤon						
SH	-[ʈa, ʈʲæ]	-kʰan	-zɔn						
RP [a]	-[ʈa, ʈʲæ]	-kʰan	-zon	-pʰaʈa, -pʰala				-ʈukuʈæ	-sɛo
BH [b]	-[ʈa, ʈʲæ]	-kʰan	-zon	-pʰala				-kuʈuʈa	-sɛo
BN [c]	-ta	-xan	-zɔn	-pʰala	-xini	-dal	-gɔ	-tukura	-heo
SCB	-ʈa, -ʈi	-kʰana, -kʰani	-ʤon						
Kamr-upi [d]	-ta	-kʰan	-zɔn	-pʰala, -pʰat	-kʰeni	-dal			
SCA [e]	-to, -ta	-kʰɒn, -kʰɒni	-zɒn			-dal			
SCO	-ʈa, -ʈi								

[a] Also for RP: /-ɖumi, -kona, -kʰona/. The latter two are probably allomorphic variants.
[b] Also for BH: /-ɖumi, -ʈʰuma, -aʃi, -gʱɔr/.
[c] Also for BN: /-silpa/.
[d] For a fuller list of the classifiers used in the Kamrupi dialect, see U. Goswami (1970:105ff).
[e] For a fuller list of the classifiers in SCA, see Kakati (1962:279ff.).

Table 5.14: Plural markers in KRDS and some other eMg lects

	NOM and OBL	Exclusively NOM	Exclusively OBL
RL	-la		
KS	-la		
MH	-la		
TH	-la		
SH	-la, gilʲa		
RP	-(gulʲæ; glʲæ)		
BH	-(gulʲæ; gilʲæ; glʲæ; la)		
BN	-gila		
SCB An.		-[e]ra	-d̪er
SCB Inan.	-gulo, -guli		
Kamrupi	-gila		
SCA	-bilak, -hɒ̃t, -bor		
SCO	-man(e), -manɔ, -kuɭa, -guɽa, -sɔbu		

There is a sharp difference in the complexity of nominal classification between western lects {RL, KS, MH, TH, SH}, and central-eastern lects {RP, BH, BN}. The former lects mark two nominal classes, *ʈa and *k^{h}an, with the subclass (or numeral classifier) *ʤɔn. The latter group of lects distinguish several more classes besides. The correlation between geographical direction and increased classificatory complexity is not accidental. From Oriya in the south-west (/ʈa/ and /ʈi/, but not /k^{h}an/), north-east through SCB and the western KRDS lects, and further north-east into the central and eastern KRDS lects as well as Kamrupi Asamiya and SCA, the complexity of classification gradually increases. The reconstruction below will make reference to this geographically conditioned complexity of declension.

Note that the additional classifiers found in RP, BH and BN also exist as independent words in these same lects, and in the other KRDS lects. This may cast doubt on their status as true classifiers, and requires further research. The morphological difference between these possible classifiers and other lexical items is that in RP, BH and BN these morphemes (e.g. /p^{h}ala/ 'strip, length') can occur directly after the noun as a suffix, whereas in the other lects the genitive case intervenes. Thus in RP, BH and BN: /bas-p^{h}ala/ 'bamboo-strip'='a/the strip of bamboo' versus in the other lects /bas-er p^{h}ala/ 'bamboo-GEN strip'='a/the strip of bamboo'. The difference between these two examples is morphosyntactic: the classifying noun has been grammaticalised within the NP, without an intervening genitive marker.

As in the illustrated system from Mahayespur, not all the classes are mutually exclusive. This can be seen in Table 5.15, where the grammatical functions of some classifiers overlap, enabling more than one classifier to be used with the one noun, for example 'bamboo'. Deeper synchronic study is required before we can say to what extent the use of each of these classifiers is grammatically as against semantically conditioned. It is quite likely that, as in the case of MH /ʈa/ and /ʤʌn/, there is some grammatical hierarchy to these classifiers in the more complex systems such as BH and BN.

Fifteen nominal concepts were elicited for specificity-classification at all eight sites. The spread of these fifteen nouns across noun classes is given in Table 5.15.

Table 5.15: Membership of classes exemplified by fifteen nouns

	Classification of nouns												
	-ʈa,-ʈə,-ʈʌ,-ta, -ɖa,-ɖʌ,-ɖə	-kʰan, -xan	-gɔ	-dal	-sɛo, -heo	-silpa	-aʃi	-gʱɔr	-d̪umi	-pʰata, -pʰala	-kuʈura, -tukura	-kʰini	-kʰona
RL	nose, bamboo, mango, betel leaf, pen, tree, cow, child, mother, person	hand, book, rope, hair											
KS	[a]												
MH	nose, bamboo, mango, betel leaf, pen, tree, cow, child, mother, person	hand, book, rope, hair											
TH	nose, bamboo, mango, betel leaf, pen, tree, cow, child, person, hair	hand, book, rope, betel leaf											
SH	nose, bamboo, mango, betel leaf, pen, tree, cow, child, mother, person, hand, hair	book, rope											

Table 5.15: continued

	Classification of nouns												
	-ʈa,-ʈə,-ʈʌ,-ta, -ɖa,-ɖʌ,-ɖə	-kʰan, -xan	-gɔ	-dal	-sɛo, -heo	-silpa	-aʃi	-gʱɔr	-d̪umi	-pʰata, -pʰala	-kuʈura, -tukura	-kʰini	-kʰona
RP	nose, bamboo, mango, book, pen, tree, cow, child, person	book, rope, hand			bamboo				bamboo		bamboo mango		betel-leaf, child, person
BH	nose, mango, hand, book, pen, tree, cow, child, person	book, rope, hair, betel leaf, tree, children, person			rope		mango	mother	bamboo	bamboo, mango, betel leaf	bamboo mango		
BN	pen, cow, person, child, mother, hair	hand, book, betel leaf	nose, mango, pen, cow, mother, child, person	hair, bamboo, rope, tree	rope	mango				mango	bamboo rope	hair	

5.4.2 Historical reconstruction of increased classificatory complexity

As noted in the synchronic overview, complexity of nominal classification increases towards the east of KRDS and in Asamiya. The complexity is not 'original'—that is, not inherited from proto Kamta—but has developed incrementally through incorporating more nouns within the grammatical set of postnominal markers of specificity. Taking what is basic across KRDS, I reconstruct the specifiers *-ʈa, *-kʰan and *-ʤɔn as inherited from the proto Kamta historical stage and probably further back still. The introduction of the classifier *-kʰan is possibly diagnostic of a common Bangla-Asamiya-KRDS propagation event, but this must be confirmed by a wider scope of comparative reconstruction. Further developments in the classification system are reconstructed as having occurred after the division of proto Kamta (AD 1550, cf. §7.3.1).

[MI 13.] Introduce as classifiers: *-pʰala, *-ʈukura, *-sɛo and assign nouns to them {RP, BH, BN}. Diagnostic value unclear.

[MI 14.] Introduce as classifier: /-ḍumi/ and assign nouns to it {RP, BH}. Nondiagnostic.

[MI 15.] Introduce as classifier: /kona;kʰona/ and assign nouns to it {RP}. Nondiagnostic.

[MI 16.] Introduce as classifiers: /-ḍumi, -ʈʰuma, -aʃi, -gʱɔr/ and assign nouns to them {BH}. Nondiagnostic.

[MI 17.] Introduce as classifiers: /-xini, -gɔ, -silpa, -dal/ and assign nouns to them {BN}. Nondiagnostic.

The nouns that become classifiers are already used in phrasal specification in other lects (e.g. /am-ɛr pʰala/ 'the strips of mango' in SH). The grammaticalisation of these nouns as classifiers seems to be conditioned by contact relations with Tibeto-Burman lects (cf. Masica 1991:250), and is thus not diagnostic of propagation events because of the possibility of independent grammaticalisation in different areas. It is nonetheless interesting that the same nouns appear to have been grammaticalised in RP, BH and BN: *-pʰala, *-ʈukura, *-sɛo. At present doubt remains over the morphosyntactic status of *-pʰala, *-ʈukura, *-sɛo as nominal classifiers in RP, BH and BN; and while this is the case, [MI 13.] cannot be considered as diagnosing PES. Further synchronic analysis of the nominal morphology of RP, BH and BN is needed in order to verify the change for these three lects; only then will it be clear whether they share classifier morphology that is innovative within KRDS.

5.4.3 History of the plural markers

The OIA plural markers were eroded during MIA, and from the start of the NIA period nouns of multitude were used as suffixes to denote plurality:

> In Assamese as in Bengali … the plural affix of O.I.A. *-ă, -ā* nouns, *-āḥ* > *M.I.A. -ā*, was reduced to *-ă* in [Apabhramsa] and lost its Pl. force … New Pl. forms had to be built up by adding nouns of multitude (Kakati 1962:93).

The forms across KRDS are reflexes of proto Kamta *-guɭa. This in turn is a reflex of the semi-Tatsama form *kula* 'herd, troop' (see Turner 1962–66; headword id. 3330). Subsequent to the grammaticalisation of this noun as a plural suffix, its form has been reduced in some KRDS lects: *-guɭa > -gula > -gla > -la. Cognates of *kula* are also found

in Bangla /gulo/, Kamrupi Asamiya /gila/, and Oriya nouns of multitude /kuɭa, guɽa/ (cf. Misra 1975:54 for Oriya). The change in vowel in Kamrupi /gila/ < *kula* also occurs in some KRDS lects. This change seems to reflect a stage intermediary between *-guɭa > -gula (> *-gVla) > -gla. During this intermediary stage the vowel—written V—became very short. The phonetic qualities of this reduced vowel were reinterpreted phonemically as /i/, rather than /u/.

Masica (1991:229) notes that 'Bengali *gulo/guli* … may be related not only to Western Assamese (Kamrupi) *gila*, but possibly also to Gawarbati *gila*, Khowar *gini*, etc. in the far northwest.' Probably all that can be said is that the semi-Tatsama form *kula* formed part of the proto Magadhan (and earlier) inheritance as one of a number of nouns that had some plural function. This form was then regularised with plural function in the lects mentioned above. This scenario of grammaticalisation is unlikely to be diagnostic of a common propagation event between Gawarbati, Khowar, Kamrupi, Oriya, Bangla, KRDS, etc.

The form of the plural marker has undergone reduction in several KRDS lects. The reduction of *-gula 'PL' > *-gla 'PL' seems to be connected to the phonological processes described in §4.4.6. Further reduction of *-gla > /-la/ in {KS, RL, MH, TH, SH, variably in RP} must be accounted for by a change specific to this morpheme:

[MI 18.] *-gla 'PL' > /-la/ 'PL' {KS, RL, MH, TH, SH, variably in RP}. Nondiagnostic.

[MI 18.] is not diagnostic of a PE because it involves loss rather than addition of linguistic material, and there is a strong phonotactic motivation for this loss. Sequences of CCC are dispreferred in KRDS, and the suffix -gla creates such sequences when attached to a consonant final stem, for example /kitapgla/ 'books'.

The history of the SCB plural for animate nouns /-[e]ra/ has already been given above under §5.3.3.

5.5 Personal pronoun systems: description

The personal pronoun systems are presented in this section with only minimal comments on peculiar contrasts and forms. Detailed reconstruction of the proto Kamta pronominal system comes in §5.6 after all the systems have been individually sketched.

5.5.1 Kishanganj (KS)

The pronoun system collected in a village area outside Kishanganj town, and reported for other southern areas of Kishanganj district of Bihar, is given in Table 5.16. Empty cells in this and later pronominal tables indicate categories identified as ungrammatical by the informant(s). 'Oblique' pronouns take case suffixes to indicate their function within the clause. The 'nominative', or direct pronouns take no suffix, and function as subject in a clause. Underscore marks indicate the position of the case marker in an oblique form which is also suffixed by -ʌ. For example: kəha-_=o → /kəha**k**o/ 'whomever' = 'INT–DAT=INDF'.

Table 5.16: Kishanganj system of personal pronouns

Person	SG.NOM	SG.OBL	PL.NOM	PL.OBL
1	mu˜i	mo-	hamra	həmsa-; həmtʃa-
2	t̪u˜i	t̪o-	t̪umra	t̪umsa-
3.PROX	jə˜haj	jəha-	ɛra; ɛmra	isma-; isa-
3.DIST	wəha˜j	wəha-	wora; ʌmra	usma-; usa-
INT.DEF	kɛ	kəha-		
INT.INDF	kah=o	kəha-__=o(b^{h}i)		
REL.DEF	ʤaj	ʤəha-		
REL.INDF	ʤah=o			

The KS system differs from RL, but is similar to the other KRDS sites in only distinguishing singular and plural pronouns. The plural forms may also be used for high singular functions. In KS the on-glide to the third person singular pronouns (variable in RL below) is more categorically established, for example: /jə˜haj/ 'PROX' and /wəha˜j/ 'DIST'. Obliqueness in the plural pronouns is signalled by a morpheme /t̪s, s/, which appears in the first and second person pronouns in a different position compared with the third person pronouns. Reconstruction of the historical morphology of this morpheme is attempted in §5.6.1. Nominative plurality is signalled by the element /ra/, also considered in §5.6.1.

5.5.2 Rangeli (RL)

The pronoun system reported for Rangeli and other areas of Morang district of Nepal, is given in Table 5.17.

Table 5.17: Rangeli system of personal pronouns

Person	SG.L.NOM	SG.L.OBL	SG.H	PL
1	mu˜i	mo- [a]	hama	hama-la
2	t̪u˜i	t̪o-	t̪am^{ɳ}a; t̪ʌm^{ɳ}a	t̪am^{ɳ}a-la; t̪ʌm^{ɳ}a-la
3.PROX	jɛ˜	jɛha-	jɛm^{ɳ}a; ɛm^{ɳ}a	jɛm^{ɳ}a-la
3.DIST	wa˜hɛ˜ [b]	waha-	am^{ɳ}a; ʌm^{ɳ}a	am^{ɳ}a-la; ʌm^{ɳ}a-la
INT.DEF	kəhaj	kaha-		kahaj-la
INT.INDF	kah=ʌ	kaha-__=ʌ		
REL.DEF	ʤəhaj	ʤəha-		ʤahaj-la
REL.INDF	ʤah=ʌ	ʤaha-__=ʌ		

[a] Wilde (2008:86) gives three oblique forms for the singular pronouns in Nepal Rajbanshi: /mo-/ with GEN and DAT case; /mʌ-/ with inclusive clitic /=hʌ/; /mʌhʌ-/ with case and the emphatic or inclusive clitic.

[b] As noted in Wilde (ibid.), the pronunciation of this morpheme varies considerably in Nepal Rajbanshi.

The Rangeli system distinguishes three persons across the number categories of Low Singular, High Singular and Plural. This is peculiar among the eight KRDS lects examined here, and is interpreted diachronically in §5.6. Speakers report that SG.L forms are used in casual conversation, but that in formal situations, such as with one's father-in-law, 'we

don't speak **mui tui**, we speak **hama tama**'. That is, SG.H forms are used in formal conversation styles, and SG.L in casual conversation styles. A thorough sociolinguistic study of the use of these different low and high pronouns remains to be done. The use of inherited plural pronouns as high singular has been accompanied by the innovation of new plural forms, extended by the plural suffix /-la/ for example: /mui/ '1SG.L', /hama/ '1SG.H', /hama-la/ 'we'.

Third person pronouns are deictics which distinguish distal (far) and proximal (near) positions against the categories low singular, high singular, and plural.

In all KRDS lects, indefinite pronouns are formed by attaching the associative clitic /=ʌ;ɔ;o/ 'even, also' to the interrogative pronoun. For example, in RL the interrogative pronoun is /kəhaj/ 'who?', and the indefinite pronoun is /kah=ʌ/ 'someone, anyone, whoever'. (In some lects including RL the attachment of the clitic alters the rhythm of the word with some effect on the preceding vowels.)

We may note in RL the phonetic variation between onglided and pure vowels in the opening syllables of the third person pronouns, for example: /jɛmʱa; ɛmʱa / '3SG.H' = 'this respected one here'. There is variation also in the constituent phonemes of the second and third person high singular and plural pronouns, for example: /t̪amʱa-la; t̪ʌmʱa-la/ '2SG.H', /(amʱa-la; ʌmʱa-la/ '3SG.H'.

The Rangeli (RL) system, like the Mahayespur (MH) system (§5.5.3), distinguishes nominative and oblique forms only in the singular pronouns, with the distinction neutralised in the plural pronouns (see further §5.6).

5.5.3 Mahayespur (MH)

The pronoun system collected in Mahayespur, and reported for other areas of eastern Jhapa district of Nepal and southern Darjeeling district of West Bengal, is given in Table 5.18.

Table 5.18: Mahayespur system of personal pronouns

Person	SG.NOM	SG.OBL	PL.NOM	PL.OBL
1	mu˜i	mo-	hama	hama
2	t̪u˜i	t̪o-	t̪ʌmʱa	t̪ʌmʱa
3.PROX	ɛ	ɛ-	ɛmʱa	ɛmʱa
3.DIST	ʌj	ʌ-	ʌmʱa	ʌmʱa
INT.DEF	kaj	kəha-	kaɛ kaɛ	kəha-__ kəha-
INT.INDF	kah=ʌ	kəha-__=ʌ	kahʌ kahʌ	kəha-__=ʌ kəha-__=ʌ
REL.DEF	dzahɛ	dzəha-	dzahɛ dzahɛ	dzəha-__ dzəh-a-
REL.INDF	dzah=ʌ	dzəha-__=ʌ	dzahʌ dzahʌ	dzəha-__=ʌ dzəha-__=ʌ

In the interrogative and relative pronouns of this set plural number is indicated by doubling of the singular forms. Analysis of oral texts is needed to find out how prominent this plural marking strategy is in actual use. This strategy may also be present in KS and RL, but failed to show up in the data elicited for this study. (See Wilde [2008:101, 102] who analyses the function of this reduplication strategy as individualised plurality.)

In MH there is no difference in form between the nominative and oblique-base plural pronouns, except in the interrogative and relative pronouns.

5.5.4 Thakurgaon (TH)

The pronoun system collected in a village near Thakurgaon town of Bangladesh is given in Table 5.19. For some categories there are contrasting forms reported for the local Muslims and for the local Polia/Rajbanshi Hindus. Forms reported for Muslims are indicated by {M}, and for Hindus by {H}. The data were collected with a Muslim speaker and his Hindu and Muslim friends, and the variation is confirmed by the dialectological data also collected during this study (Appendix D of Toulmin 2006).

Table 5.19: Thakurgaon system of personal pronouns

Person	SG.NOM	SG.OBL	PL.NOM	PL.OBL
1	mui	mo-	hamra	hama-
2	t̪ui	t̪o-	t̪umrɦa	t̪umɦa-
3.PROX	iaˇ	iaˇ-	imrɦa	imɦa-
3.DIST	uaˇ	uaˇ-	umrɦa	umɦa
INT.DEF	ke {M}, kaj{H}	ka-	kela {M}, kara {H}	ka-__ ka-
INT.INDF	keh=o	ka-__=o		ka-__=o ka-__=o
REL.DEF	dzaj	dza-	dzejla	dza-__ dza-
REL.INDF	dze keh=o			

This system is like MH in utilising reduplication for plurality in the interrogative and relative pronouns. The combination of aspirated nasals in the plural oblique forms (e.g. /t̪umɦa/ '2PL.OBL') and aspirated rhotics in the plural nominative forms (e.g. /t̪umrɦa/ '2PL.NOM') is unique among the eight KRDS sites and significant for the reconstruction in §5.6.

The interrogative PL.NOM form /kela/ reported for Muslims is a newer plural than the Hindu equivalent /kara/. The Muslim variant results from agglutination of the INT.SG.NOM pronoun /ke/ with the productive plural morpheme /-la/ (cf. §5.4).

5.5.5 Shalkumar (SH)

The pronoun system collected with speakers of Shalkumar, in central Jalpaiguri district, West Bengal, is given in Table 5.20.

Table 5.20: Shalkumar system of personal pronouns

Person	SG.NOM	SG.OBL	PL.NOM	PL.OBL
1	Mui; moj	mo-	ham(e)ra	hama-
2	t̪ui	t̪o-	t̪ɔmora; t̪omra	t̪oma(-__-la)-
3.PROX	ijai	ija-	imirʲa	ima-
3.DIST	uwai	uwa-	umurʷa, umra	uma-
INT.DEF	kaj	ka-	kajgula	kunla-; ka-__ ka-
INT.INDF	kah=o	kaho-__=o	kah=o kah=o	
REL.DEF	dzaj	dza-	dzaj dzaj, dzeila	
REL.INDF	dze kah=o			

Peculiar to the Shalkumar data is the insertion of a vowel between sonorant consonant clustersThe particular vowel that is inserted is determined by the preceding vowel, thus /ham(e)ra/, /t̪ɔmora/, /imirʲa/, and /umurʷa/ (cf. §4.4.8). The pattern of pronominals is otherwise highly similar to RP and BH below (minus the innovative relative plural form of Rangpur).

5.5.6 Rangpur (RP)

The pronoun system collected with speakers in and around Rangpur town, Bangladesh, is given in Table 5.21.

Table 5.21: Rangpur system of personal pronouns

Person	SG.NOM	SG.OBL	PL.NOM	PL.OBL
1	mu˜i; mo˜i	mo-	hamra	hama-
2	t̪u˜i; t̪o˜i	t̪o-	t̪omra	t̪oma-
3.PROX	æ˜i	iæ-	emra	ema-; imæ-
3.DIST	ɔ˜i; t̪a˜i	uæ-	omra; t̪amra	oma-
INT.DEF	kãi	ka-		–
INT.INDF	ka˜jo	ka-__=o		–
REL.DEF	za˜i	za-	zamra (DIST), zemra (PROX)	zama-
REL.INDF	za˜jo	za-__=o		–

Three features of this system warrant some comment for their variance from the broader KRDS pattern. Firstly, this is the only system to have extended the plural nominative and oblique elements /-mra/ and /-ma/ to the relative or subordinating conjunctions: /zamra, zemra, zama-/. Interestingly, these new relative pronouns further distinguish the categories distal versus proximal. Secondly, along with standard third person distal pronouns in /ɔ-;o-/, the data also include an anaphoric form (listed under 3.DIST) in /t̪a-/. Functionally equivalent forms exist in the other central and eastern KRDS lects, but not in RL, MH, or KS to my knowledge. Thirdly, the variation [mu˜i; mo˜i] and [t̪u˜i; t̪o˜i] is part of a confusing historical picture of the development of these pronominal forms, see further §5.6.1.

5.5.7 Bhatibari (BH)

The pronoun system collected with speakers from villages around Bhatibari—on the border between south-eastern Jalpaiguri district and north-eastern Cooch Behar district, West Bengal—is given in Table 5.22.

Table 5.22: Bhatibari system of personal pronouns

Person	SG.NOM	SG.OBL	PL.NOM	PL.OBL
1	mui	mo-	(h)amra	(h)ama-
2	t̪ui	t̪o-	t̪omra	t̪oma-
3.PROX	iŋʲɛj; ijɛj	iŋʲɛ-; ijɛ-	ɛmra, imrʲɛ	imʲɛ-
3.DIST	uŋʲɛj; uwɛj	uŋʲɛ-; ujɛ-	umrʲɛ	umʲɛ-
INT.DEF	kaj	ka-	kaj kaj	ka-__ ka-__
INT.INDF	kaŋ=o, kaj=o	kaŋ-__=o	kaŋ=o kaŋ=o	kaŋ-__=o kaŋ-__=o
REL.DEF	dzaj	dza-	dzaj dzaj	dza-__ dza-
REL.INDF	dzaŋo			

This pronominal system is substantially the same as for Rangpur above, though with reduplication as a strategy for marking plurality in the place of Rangpur's innovative relative plural forms.

5.5.8 Bongaigaon (BN)

The pronoun system collected with speakers of Bongaigaon lect, in western Assam is given in Table 5.23.

Table 5.23: Bongaigaon system of personal pronouns

Person	SG.NOM	SG.OBL	PL.NOM	PL.OBL	H.NOM	H.OBL
1	mɔj	mɔ-	ami; amira	ama-		
2	tɔj	tɔ-	tumi; tumira	toma-; tuma-	apuni	apona-
3.PROX	ɛ	e-	imira	ima-		
3.DIST	oj, hi	ta-	umira, tamira	uma-, tama-		
INT.DEF	kaj	ka-	kaj kaj	ka-__ ka-		
INT.INDF	kabaj	kaba-	kabaj kabaj	kaba-__ kaba-		
REL.DEF	(d)zaj	(d)za-	(d)zigila	(d)za-__ (d)za-		
REL.INDF	zabaj	zaba-				

Several aspects of this pronominal system diverge from the other KRDS systems. The first and second person singular pronouns /mɔj, tɔj/ are almost identical with SCA.[15] The second person pronouns include a high honorific form, which lacks cognates elsewhere in KRDS, but has cognate forms in SCA as well as SCB (see §5.6.3). Thirdly, the third person singular distal 's/he there' is /oj/ or /hi/. The latter may have an anaphoric function. Lastly, the indefinite pronouns based on /kaba-/ are markedly different to the equivalent forms in other KRDS lects, though similar to SCA.

[15] SCA forms are /mɒe, tɒe/ '1SG, 2SG'. SCA has four distinct phonemes in the back vowels /ɒ, ɔ, o, u/, to BN's three /ɔ, o, u/.

5.5.9 Standard Colloquial Bangla (SCB)

For the sake of comparison, and given the influence of the regional Standard languages on certain KRDS lects, the pronominal systems of SCB and SCA are also outlined briefly. The SCB system given in Table 5.24 is based on P. Dasgupta (2003:367) and T. Bhattacharya (2001:68). The levels of honour are given in the leftmost column: L (low), NT (neutral), H (high honour). Anaphoric function /ʃɛ/ is distinguished from distal /o/ in the third person.

Table 5.24: SCB system of personal pronouns

Person	SG.NOM	SG.OBL	PL.NOM	PL.OBL
1	ami	ama-	amra	amaḍer
2.L	ṭui	ṭo-	ṭora	ṭoḍer
2.NT	ṭumi	ṭoma-	ṭomra	ṭomaḍer
2.H	apni	apna-	apnara	apnaḍer
3.NT.PROX	e	e-	era	eḍer
3.H.PROX	ini	e˜-	e˜ra	e˜der
3.NT.DIST	o	o-	ora	oḍer
3.H.DIST	uni	o˜ra	o˜ra	o˜ḍer
3.NT.ANP	ʃe		ṭara	
3.H.ANP	ṭini		ṭa˜ra	
INT.DEF	ke	ka-		
INDF	ke=o	kau-		
REL.DEF	ʤe	ʤa-		
REL.INDF	ʤe ke=o	ʤe kau-		

The SCB pronoun system is considerably more complicated than the KRDS systems, due to the categorisation for honour in second and third person. While in KRDS there are generally no special forms to distinguish high and low honour, SCB distinguishes three levels of honour in the second person forms /ṭui, ṭumi, apni/, and two levels in the third person forms /o, uni/, /e, ini/, etc. In the second person, KRDS generally only has /ṭui/ for singular number, and the further option of using the plural pronoun /ṭomɦa, ṭomra/ in cases of high honour. As a result, KRDS /ṭui/ is not functionally equivalent to SCB /ṭui/. Rather, the function of KRDS /ṭui/ is equivalent to the functions covered by both SCB /ṭui/ *and* /ṭumi/. The functional equivalent of SCB /apni/ is KRDS /ṭomɦa; ṭomra/, but this KRDS pronoun also includes plural function, so the functional equivalence is not one-to-one. Similar differences in the categorisation of honour exist between the third person pronouns of SCB and KRDS.

These structural differences lead to misunderstandings, largely on the part of SCB speakers, who mistakenly assume that KRDS /ṭui/ exists in the same structural relations of honour as SCB /ṭui/ when in fact the structure of the systems is quite different. I have had Bengali mother tongue speakers say to me ‘Rajbanshis are rude, they use *tui* when they speak to me’. This is a misunderstanding of the functional relations of /ṭui/ ‘2SG’ within the pronominal system of KRDS lects.

Note also that the function of the SCB pronominal element /-ra/ differs from the function of KRDS /-ra/. In KRDS and other eastern Magadhan lects the morpheme is

restricted to the pronominal system, while in SCB it is a general marker of plurality for animate nouns (see §5.3.3).

5.5.10 Standard Colloquial Asamiya (SCA)

The structure of the SCA pronominal system is similar to SCB in its categorisation of honour in second and third persons. However, there are several other differences with SCB and with KRDS, as can be seen in the data in Table 5.25. The data are from Goswami and Tamuli (2003) and Kakati (1962).

Table 5.25: SCA system of personal pronouns

Person	SG.NOM	SG.OBL	PL.NOM	PL.OBL
1	mɒj	mo-	ami	ama-[16]
2.L	tɒj	to-	tɒhɒ̃t	
2.NT	tumi	toma-	tomalok	
2.H	apuni	apona-	aponalok	
3.NT.PROX:M	i	ia-	ihɒ̃t	
3.NT.PROX:F	ei		xihɒ̃t	
3.H.PROX	eo̜, ek^{h}et		eo̜lok, ek^{h}etxɒkɒl	
3.NT.DIST:M	xi	ta-	xihɒ̃t	
3.NT.DIST:F	tai			
3.H.DIST	teo̜, tek^{h}et		teõlok, tek^{h}etxɒkɒl	
INT.DEF	kon	ka-		
INT.INDF	konoba	karoba		
REL.DEF	zi	za-		

SCA stands out in eastern Magadhan for its categorisation of gender in the third person. The distinction is maintained only in the nominative pronouns (/i/ ‘he’, /ei/ ‘she’) and not in the oblique ones (e.g. /ia-r/ ‘his, her’). The gender distinction is neutralised for the third person pronouns with high honour, for example: /eo̜, ek^{h}et/.

The plural element /-ra/ mentioned above for SCB and KRDS is noticably absent from SCA (though it was present during early Asamiya, cf. §5.3.3). Plurality is marked either through use of different lexemes /mɒj/ ‘I’ versus /ami/ ‘we’, or by the suffixes /-hɒ̃t/ or /-lok/. The suffix /-hɒ̃t/ is always applied to low honour pronouns, and sometimes to neutral honour pronouns, but never to high honour pronouns. The suffix /-lok/ is high honour, and sometimes neutral honour.

5.6 Personal pronoun systems: reconstruction

Having sketched the contemporary pronominal systems of KRDS and its influential neighbours SCB and SCA, the present section reconstructs the historical change events that

[16] Data missing for the remainder of this column.

derived the contemporary systems from earlier ones. Structurally-general innovations that apply across person categories are reconstructed first in §5.6.1, followed by a blow-by-blow reconstruction of forms in each category of person (§5.6.2–§5.6.4), as well as in the interrogative (§5.6.5) and relative (§5.6.6) pronominals. The proto Kamta pronoun system that results from the reconstruction is presented here in advance, in order to aid the reader in following the discussion below.

Table 5.26: Reconstructed personal pronouns for pKmt

Person	SG.NOM/INS	SG.OBL	PL.NOM	PL.OBL
1	*mui	*mo-	*hamra	*hama-[ʃa-]
2	*t̪ui	*t̪o-	*t̪omⁿra	*t̪omⁿa-[ʃa-]
3.PROX	*[ɛ˜j; i˜haj]	*i˜ha-	*[ɛmⁿra; imⁿra]	*[ɛmⁿa-; imⁿa-, ɛʃma-; iʃma-]
3.DIST	*[o˜j; u˜haj]	*u˜ha-	*[omⁿra; umⁿra]	*[omⁿa-; umⁿa-, oʃma-; uʃma-]
INT	*kahɛ	*kaha-		
REL	*dzahɛ	*dzaha-		

5.6.1 General structural changes in personal pronouns

The KRDS lects are treated in the same order as above, with departures from this order only when it is necessary to describe innovations across several lects.

The structure of the RL system differs from the other KRDS sites by formally distinguishing low singular pronouns (e.g. /mu˜i/ ‘1SG.L’) from high singular (e.g. /hama/ ‘2SG.H’) and general plural (e.g. /hamala/ ‘1PL’). Important points to note are: high and low are only distinguished in the singular number, and are distinguished across all three persons. This system of honour marking is completely different from the SCB and SCA systems which distinguish three levels of honour across both single and plural pronouns, but only in the second and third persons.

The RL system of honour marking is not an inherited feature of proto Kamta, but a recent innovation through the shift in meaning of the inherited plural > high singular. New plural pronouns have been formed by suffixing the inherited plural pronouns with the ending /-la/ ‘PL’ (cf. §5.4.3). Notably, given RL’s Hindi and Bihari language contact, a similar shift also occurs in varieties of those languages whereby the old plural /həm/ ‘we’ functions as a singular pronoun ‘I’ in the place of inherited /məi˜/ ‘I’. The new plural is formed by a help word or suffix, such as /log/ ‘people’ in some varieties of Hindi, for example /həm log/ ‘we’.

The inherited proto Kamta pronoun system is reconstructed as distinguishing three persons, with singular and plural number, in nominative and oblique functions, but without pronominalised honour marking. The RL system diverges from this reconstruction by the following changes:

[MI 19.] pKmt pronouns with PL function > SG.H function {RL}. Nondiagnostic.

[MI 20.] pKmt pronouns inherited with PL function are suffixed by /-la/ ‘PL’ > PL function {RL}. Nondiagnostic.

The second structural divergence in pronouns occurs in the plural pronouns of RL and MH: /hama, t̪ʌmʱa/ etc. These pronouns diverge from the general KRDS pattern by not employing distinct pronominal forms for nominative and oblique arguments. The simplest historical solution is to reconstruct the merger of nominative and oblique categories in the MH and RL plural pronouns, with retention elsewhere in KRDS:

[MI 21.] Pronouns with function PL.OBL are extended to general plural function (thus including PL.NOM) {RL, MH}. Diagnostic.

This change is diagnostic of a PE, because the merger of these morphological categories is ecologically distinctive. The two lects are also adjacent to one another, and hence the range of propagation is sociohistorically plausible.

The element /-ra/ 'PL.NOM' has been reconstructed in §5.3.3 as an inherited feature of the proto Magadhan pronominal system.

In Kishanganj, where the nominative and oblique distinction is maintained, plural oblique is marked in a peculiar manner. The typical KRDS marking of oblique pronouns is with the suffix /-a/, for example /ham-**ra**/ '1PL-PL.NOM', /ham-**a**-/ '1PL-OBL'. However, in Kishanganj obliqueness is marked on plural pronouns by an *-s-* element (with a variant allomorph /ʧ/), in addition to /-a/. This element comes between the pronominal base and the typical oblique suffix /-a/ in both first person and second person pronouns: /ham**s**a-; ham**ʧ**a-/ and /t̪um**s**a-/. The same element occurs in the third person plural oblique pronoun 'them': /i**s**ma-; i**s**a/ and /u**s**ma-; u**s**a/. However, in these forms the *-s-* element precedes a variable *-m-* element.

Two questions must be answered in order to reconstruct the history of these forms: Is the element /s/ a proto Kamta retention or a post-proto Kamta innovation? And why does it occur *before* the variable *-m-* in the third person, rather than after it (as in the first and second person pronouns)? Both questions must be answered perspicuously by any proposed etymology of the *-s-* element. Kakati (1962:295) presents some pertinent data from early Asamiya:

> -s- . Used in E.As. [Early Asamiya] only after oblique forms of the pronouns of the first and second persons (*āmāsā-k*, to us; *āmāsā-r*, of us; *tomāsā-t*, in you). It is found also in Bengali (Siripuria, Purneā), *hams-ār*, our; *tums-ār*, your (L.S.I. [Linguistic Survey of India], vol.I, p.354). In this connection cf. Bihārī (Bhoj-puri) *-sa* (*ghoṛā-sa*; horses) (L.S.I., vol.II, p.224)
> The origin of this -sa- seems obscure. Dr. Chatterji taking the Assamese dative form in *-sāk* (*āmāsā-k*; *tomāsā-k*) alone, affiliates *-sāk* to inscriptional *-sat-ka-* (O.D.B.L. 504).
> [The term 'Siripuria' and its classification as 'Bengali' are from the LSI. Today's speakers in KS use the term 'Surjapuri' and do not class it as Bengali—MT]).

Unfortunately Kakati does not name the early Asamiya document in which he finds the relevant forms with /-s-/. The written documents categorised as early Asamiya are at points closely connected with early KRDS stages of development. Kakati sets the time frame for early Asamiya as 'from the fourteenth to the end of the sixteenth century' (ibid.:13). During this period Asamiya literature was written under the patronage of the Koch Kings of Kamatapur and in the Kamrupi or western Asamiya dialect. This was also the period when proto Kamta features were innovated (cf. §7.3.1). It is not out of the question that mixing of some proto Kamta features may have occurred in the early Asamiya document he mentions, though this hypothesis remains to be checked, and the rest of the argument below does not depend on it.

The evidence from early Asamiya suggests that the *-s-* element has been part of the linguistic history of the area for some time, and should be considered an inheritance from the proto Kamta stage rather than a KS innovation. However, other etymologies for this element are possible besides that put forward by Chatterji (which Kakati termed 'obscure'). The two crucial pieces of data are the following:

(1) in KS the *-s-* element occurs not only in the first and second plural pronouns, but also in the third person plural pronouns before a variable *-m-* element,

(2) in early Asamiya Kakati (1962:295–296) records an alternative plural marking strategy in the pronouns:

> *-saba, samba*: (OIA *sarva* > MIA *sabba, *samba*). Used in [early Asamiya] as Plural suffixes after oblique forms of the second and third person pronouns; e.g. *tomā-sab*, you all; *tā-sambār*, of them all.

Taking all this evidence together, it seems quite likely that the *-s-* element in KS (and early Asamiya) is from MIA *sabba, *samba* 'all', rather than from *-satka*. Chatterji after all had made the latter reconstruction based on the mistaken identification of the whole element *-sāk* as a plural oblique ending. However, as Kakati points out, the *-k* element does not code general oblique but is specifically dative, and it is *-sā* alone which codes plural oblique function. Kakati shows this by citing forms with other (non-dative) case endings, for example *āmāsā-r* 'our'.

If we take the origin of the *-s-* element in KS to be MIA *sabba* > **samba* 'all' this accounts for the two pieces of data presented above:

(1) The first and second person plural oblique pronouns /hamsa-, t̪omsa-/ are reflexes of *ham- and *t̪om- suffixed with *-a. These oblique bases were supplemented at an early stage by a plural word *ʃɔm 'all' < *ʃɔmbɔ. The historical changes were something like: *ham-a ʃɔm *ham-a-ʃm-a- > *hamaʃa > *hamʃa- '1PL.OBL'. The reduction of *ʃm > *ʃ is phonologically rather than morphologically conditioned—cf. the variation of the nasal in KS /is(m)a-/ 'them'. The third person pronoun retained the nasal longer *iʃɔm-a > /isma-/ > /is(m)a-/

(2) This hypothesis also accords with the use of derivatives of MIA *sabba, *samba* 'all' in early Asamiya for plural oblique function, for example *tā-sambā-r* 'of them all'.

This proposed etymology is perspicuous in explaining the position of the *-s-* element in KS and early Asamiya. Following this reconstruction, key stages leading to the proto Kamta plural oblique pronouns are given as follows:

Table 5.27: Reconstructed changes in plural oblique pronouns

	1PL.OBL		2PL.OBL		3.OBL (DIST)	
pMg	*amⁿa-		*t̪omⁿa-		*o-	
pKmr	*amⁿa-	*hama ʃɔmba-	*t̪omⁿa-	*t̪omⁿa ʃɔmba-	*o ʃɔmba-	
pKmt	*hama-	> *hama-ʃma-	*t̪omⁿa-	> *t̪omⁿa-ʃma-	> *o ʃɔma-	*omⁿa-
	*hama-	> *hamaʃa-	*t̪omⁿa-	> *t̪omaʃa-	> *oʃma-	*omⁿa-

In accordance with the hypothetical sequencing shown in Table 5.27, [MI 22.] is tentatively reconstructed as part of the proto Kamrupa stage—ancestral to both proto Kamta (KRDS) and proto-eastern Kamrupa (Asamiya) (cf. §7.3.4). A similar extension of pronouns with a cognate morpheme is found in Maithili, so [MI 22.] is not unique to Asamiya and KRDS linguistic history. Therefore this change will not be considered diagnostic of a propagation event until the relations with Maithili are better understood. Also let it be noted that the proto Kamrupa stage is not yet well established by diagnostic changes (cf. §7.3.4), and thus further studies may revisit the hypothesis that [MI 22.] occurred during a hypothetical proto Kamrupa stage.

[MI 22.] > *ʃɔmba- 'PL.OBL' in pronoun declension {KRDS, early Asamiya} (tentatively proto Kamrupa stage). Supportive, not diagnostic.

[MI 23.] *hama-[ʃɔmba-] 'PL.OBL' > *hama-[ʃɔma-] > *hama-[ʃa-] '1PL.OBL', and the equivalent changes across the second and third person pronouns. {KRDS, ?early Asamiya}. Diagnostic value unknown.

Before moving on from the Kishanganj pronouns, there is one further feature which requires some discussion. The third person nominative plural pronouns in this lect incorporate some variation: /ɛra; ɛmra/ '3PL.PROX'; /wora; ʌmra/ '3PL.DIST'. The variable loss of the *-m-* element is unique within the KRDS area to the Kishanganj and adjacent Dinajpur areas. The fine grained dialectological data collected during the second stage of the project (see Appendix D of Toulmin 2006) show that some KRDS lects around Dinajpur have carried this phonological reduction of pronouns further still: *hamra > /hara/ 'we', *t̪omɦra > /t̪ora/ 'you PL', and *omɦra > /ora/ 'they' (see sites 56 and 57 in Appendix D of Toulmin 2006). The phonological change in KS is restricted to /wora, ʌmra/ < *omɦra and does not affect the other pronouns.

Of the KRDS pronouns systems, the BN system is most divergent from the proto Kamta system. Several of the changes affecting the BN system are general across the categories of person. Firstly, the pronouns /mɔj, tɔj/ '1SG, 2SG' are distinct from the forms which are otherwise general across KRDS: /mui, t̪ui/ '1SG, 2SG'. The /mui, t̪ui/ forms are also found across many Bangla dialects and earlier Oriya:

> **The direct form** *muˇiˇ/muˇ* < OIA instrumental singular … > *maeˇ* > *maiˇ* and on the analogy of *tu/tui mai* > *muiˇ*, and *muiˇ* > *muˇ* by shortening or due to the influence of *tu.* … Old and Middle Oriya: *muiˇ, muˇ* (Misra 1975:84, and see ibid.:87 for second person forms in Old Oriya).

The variation between *t̪ui and *t̪ɔi, *mui and *mɔi is a complex matter in eastern Magadhan history. As Misra notes, old Oriya variants (muiˇ; muˇ/ and /t̪uiˇ; t̪uˇ/ all had the high vowel /u/, and in modern Oriya the regularised forms are /muˇ, t̪uˇ/. In Asamiya the forms are /mɒj, tɒj/—with a low back vowel—and Kakati makes no statement about the corresponding forms in early Asamiya literature. Chatterji describes two variants for the first person singular instrumental pronoun in the *Caryās* (which he labels Old Bengali): *mai, moe*. The former he considers a retention from MIA, and the latter an innovative instrumental built from the oblique base *mo-* and the instrumental case marker -/e;eˇ/. For middle Bangla, Chatterji describes a high degree of variation (which may be largely orthographic, rather than phonological): '«mōē, mōēˇ, mōñaˇ, mōñē, mōñå, mōñi, muñi, muˇhi, muyiˇ, mōi », etc.' (1926:811).

While it is clear that the raised vowel in /mui/ constitutes an innovation, it is not clear which exact innovation should be reconstructed. Misra accounts for the raising in Oriya

/mɔi/ > /mui/ by analogy with an inherited second person singular pronoun /t̪u/ (see quote above). Kakati explains the same raising /ɔ/ > /u/ in Bangla as regressive vowel harmony triggered by the following /i/ (1962:312). Chatterji gives a third account of the change by proposing Bangla /t̪ui/ < /t̪ɔi/ through 'the influence of the oblique «tō-»' (1926:817). For both Misra and Chatterji, the explanation involves analogical change—either across persons (Misra), or across functions within the same person category (Chatterji). Kakati's explanation is phonological rather than morphological. All three processes are plausible, and thus the exact characteristics of the change remain an open question, with no unambiguous solution possible at the present time.

[MI 24.] MIA pronouns *mai* '1SG.INS', *tai* '2SG.INS' > NIA pronouns /mui/ '1SG.NOM', /t̪ui/ '2SG.NOM' {old Oriya, middle Bangla, KRDS}. Non-diagnostic.

This vowel raising, being consistent with both analogical and phonological pressures, is an unlikely contender for a propagation event. The possibility of independent replication is considerable. In some (or all) areas it is possible that the raising is due to regressive vowel harmony, in some areas the raising may be due to analogy and pressure to regularise across pronominal paradigms. Given the similar pronominal and phonological starting conditions across eastern Magadhan lects, it is easily conceivable that [MI 24.] should have occurred through independent replications, thus constituting more than one propagation event.

Further support for the proposal of independent replication comes when we consider the range of the change. It is highly implausible that a propagation occurred between Oriya, Bangla and KRDS (and before old Oriya at that), but *excluded* Asamiya. The existence of historical stages common to Asamiya and Bangla, as well as Asamiya and KRDS is hypothesised in Chapter 7. There is insufficient evidence, however, to warrant the reconstruction of a common stage ancestral to Bangla, Oriya and proto Kamta. Either the change [MI 24.] was independently replicated in the various areas of eastern Magadhan, or it was inherited as variation from the proto-eastern Magadhan stage with independent regularisation in some eastern Magadhan descendants.

The pronouns /mɔj, tɔj/ 'I, you' in BN are similar to Asamiya and distinct from the rest of KRDS. Recall that BN has mixed linguistic ancestry—inheriting features from both proto Asamiya and proto Kamta. The pronominal forms /mɔj, tɔj/ are part of BN's linguistic inheritance from proto Asamiya.

[MI 25.] /mɔj, tɔj/ 'I, you' {BN}. Diagnostic of contact relations with Asamiya.

Similarities between the BN and Asamiya pronoun systems also include the plural first person form (cf. §5.6.2), the high second person forms (cf. §5.6.3), the third person form /hi/ (cf. §5.6.4) and the indefinite forms in /kaba-/ (cf. §5.6.5).

This section has reconstructed the changes in KRDS pronouns which are generalisable for more than one category of person; the next sections examine changes which are specific to a single category of person.

5.6.2 First person pronouns

The first person singular nominative pronoun ('I') is reconstructed for proto Kamta as *mui (cf. Table 5.26). This pronoun has been retained in all eight of the KRDS lects with the exception of BN. The corresponding BN form is /mɔj/ '1SG.NOM', which is identical with Asamiya, and indicative of BN's historical linkage with proto Asamiya (eastern

Kamrupa) as well as proto Kamta. The first singular oblique pronoun for proto Kamta is *mo-; it is retained across KRDS.

The first person plural pronoun is reconstructed as *hamra ‘we’, with oblique counterpart: *hama-[ʃa-] ‘1PL.OBL’. The inherited nominative *hamra has been substituted with the oblique *hama in MH and RL (see [MI 21.]). The initial *h is retained in most KRDS lects, though variably lost in BH. The corresponding BN pronoun /ami(ra)/ is ambiguous between a proto Kamta and proto eastern Kamrupa inheritance. The initial *h in KRDS is innovative and distinct from Bangla and Asamiya, though common with Bihari and Hindi.

[MI 26.] *am^{ɲ}ɛ ‘we’, *am^{ɲ}a- ‘us’ {pre-proto Kamta} > *ham-ra ‘we’, *hama- ‘us’ {KRDS}. Nondiagnostic.

The proto Kamta system is reconstructed to include variation between oblique forms *hama-[ʃa-] ‘1PL.OBL’. The first variant is found in the majority of contemporary KRDS lects, but cognates of the second variant are found both in early Asamiya and the contemporary Kishanganj (KS) lect as has been discussed under §5.6.1. I hypothesise that the variation between these two forms goes back to the proto Kamta stage and that the variation was regularised after the division of proto Kamta, possibly independently in different areas. Thus, the regularisation of *hama- ‘us’ in all lects but KS is innovative, but not diagnostic of a PE.

[MI 27.] *hama- ‘us’ regularised as first person plural oblique pronoun {KRDS except KS}. Nondiagnostic.

Finally, note that the pronoun /ami/—found in Bangla meaning ‘I’ and in Asamiya meaning ‘we’—is absent in all of KRDS excepting BN. The pronoun system in BN is highly similar to that of Asamiya, and it is most plausible that /ami/ ‘we’ in BN reflects its Asamiya linguistic ancestry and not a proto Kamta inheritance.

[MI 28.] > /ami/ ‘we’ {BN}. Supportive, not diagnostic, of contact relations with Asamiya.

The presence of this form in BN supports the hypothesis of significant contact relations between BN and Asamiya (though the formal similarity with Bangla /ami/ ‘I’ means it is not *diagnostic* of those relations).

5.6.3 Second person pronouns

The second person pronouns described in §5.5.1–§5.5.8 are reconstructed in Table 5.26 as reflexes of the forms *t̪ui ‘2SG.NOM’, *t̪o- ‘2SG.OBL’, *t̪om^{ɲ}ra ‘2PL.NOM’, *[t̪om^{ɲ}a-, t̪om^{ɲ}ʃa-] ‘2PL.OBL’. The only divergence in the singular pronouns is in BN, reflecting at this point its Asamiya linguistic ancestry (see [MI 25.] under §5.6.1). Among the plural pronouns, the RP form is slightly divergent from *t̪om^{ɲ}ra > /t̪umra/. The raising is due to analogy with the singular form *t̪ui. Given the confused picture across eMg of /u/ vs. /ɔ/ and /o/ in second person pronouns (cf. §5.6.1), [MI 29.] is not diagnostic of a propagation event.

[MI 29.] *t̪om^{ɲ}ra ‘2PL’ > /t̪umra/ {RP}. Nondiagnostic.

Of the eight KRDS lects sampled, only BN has introduced a specifically honorific second person pronoun.

[MI 30.] > /apuni/ ‘2H.NOM’, /apona-/ ‘2H.OBL’ {BN}. Supportive, not diagnostic, of contact relations with Asamiya.

Cognate forms have also been introduced into Bangla and Asamiya, as well as further afield in NIA. This honorific pronoun is traced etymologically to an erstwhile reflexive pronoun /ap-/, whose use in this sense ‘is quite recent, unknown to Middle or older New Indo-Aryan … It … appears to radiate from Delhi and to be associated with urban/Muslim/ ‘Hindustani’ influence … probably in imitation in turn of ‘elegant’ Persian usage (*perhaps independently in Bengal*)’ (Masica 1992:41, my emphasis). Put in the terms of this study, Masica does not find the introduction of this honorific pronoun to be diagnostic of a propagation event linking the central Delhi region and Bengal because of the possibility of independent replication in Bengal. The change [MI 30.] in Bongaigaon is probably due to Asamiya influence (in accordance with BN’s mixed Asamiya-Kamta ancestry). However, the similarity with other NIA lects means that the change is supportive, not diagnostic of a PE. The honorific pronoun /apun-/ is clearly not to be reconstructed as part of the proto Kamta ancestry because (a) it is a recent introduction, and (b) it is not used in KRDS beyond BN, where its presence is explicable by contact relations with SCA.

5.6.4 Third person pronouns

Among the third person pronouns there are two complex matters for reconstruction. Firstly, across KRDS it is common to find variation between th person singular pronouns starting with /u(h)a-/ and /o-/. Moreover, in areas here there are sizeable populations of Muslims and Hindus, it is common to find one variant preferred by Hindus and the other variant preferred by Muslims. However, the distribution of variants is not consistent from area to area: in the north-west of Jalpaiguri district (around Oodlabari), Muslims use /ɔj; oj/ ‘s/he’=‘3SG.NOM’, and Hindus use /uaj/; further south near Shibganj of Bangladesh (site #35, see Appendix D of Toulmin 2006) the situation is exactly reversed with Muslims using /uaj/ and Hindus /ɔj/. The best explanation for this distribution of variants is that variation was inherited from the proto Kamta stage and regularised independently in different areas along social lines. Croft’s ‘first law of propagation’ (2000:176) is relevant to this differential regularisation of variation: ‘When variants are created … one variant either (i) shifts its meaning, (ii) shifts its community, or (iii) disappears’. In the case of the inherited variation of third person singular pronouns, we have examples of options (ii) in the Muslim/Hindu differentiation, and (iii) in the regularisation of one variety in one area, for example MH has /ɔj/, but RL has /wɔhaj/ < *uhaj.

The second matter for reconstruction in third person pronouns concerns the *m (or *m^{ɦ}) element found in the plural pronouns *[om^{ɦ}ra; um^{ɦ}ra] and [ɛm^{ɦ}ra; im^{ɦ}ra]. Chatterji (1926:828) writes regarding KRDS lects:

> North Bengali uses the base তাঁ (beside a fuller তানি «tāni») for the nominative; and the form [sic] তামার «tāmāra», plural তাম(া)রা «tām(ā)rā» are honorific, with «-m-» for «-n-» or «-˜h-» of other forms of Bengali—a phonetic peculiarity which characterises this dialect: as early as c. 1555 A.C., in a letter from the Kōc king Nara-Nārāyaṇa of North Bengal to the Ahom king Su-khām-phā, we find ইমারাক পাঠাইতেছি «imā-r(= ihã̄-digå-kē) paṭhāitē-chi» *I am sending them*, [the subject of this example is either plural *we are sending them* or high singular—MT])

The origin of this *-m-* element, and its uniformity across KRDS is perplex. Chatterji attempts to explain it etymologically as cognate with Bangla /n/ in third person plural pronouns. This is a possible explanation: *[o;u]w̃ra > *[o;u]mra, but as there are no other instances in the data where we reconstruct the cluster *w̃r the phonological regularity of this hypothetical change cannot be tested at present.

As argued under §5.6.1, there is reason to reconstruct oblique plural pronouns *[oʃma-; uʃma-] with the postposed *ʃm < *ʃɔm < *samba* 'all'. It is possible that the *m element in the corresponding nominative form *[o;u]**m**ra was introduced by analogy with the *m of the oblique *-ʃm. However, there is no evidence in KRDS (outside of BN, by Asamiya influence cf. §4.3.13) for *ʃ > h. Lacking corroboration in the reconstructed KRDS phonological changes, a different explanation should be sought.

A third, and more plausible explanation is found in analogy across persons. First and second person plural forms are reconstructed as *hamra and *t̪omⁿra, flanked by the elements *mra and *mⁿra. It is quite conceivable that the similarity in these forms was reinterpreted as 'plural nominative' and extended to the third person to give *[omⁿra; umⁿra]. Similarly, in the oblique pronouns the first and second persons are *hama-, *t̪omⁿa- and it is conceivable that the nasal stop element *mⁿa was analogically extended to the third person to give: *[omⁿa-; umⁿa-]. These reconstructed changes of morphological reinterpretation and analogical extension are further supported by the *relative pronouns in RP which have also incorporated the morphological elements* /-mra, -ma/ to give /zamra/ 'REL.NOM.PL' and /zama-/ 'REL.OBL.PL'(see §5.6.6 below).

[MI 31.] *mⁿra reinterpreted as 'PL.**NOM**' in pronoun system, and extended as such to third person *[o;u]mra {KRDS, also some Hajong lects}. Diagnostic.

[MI 32.] *mⁿa- reinterpreted as 'PL.**OBL**' in pronoun system, and extended to third person *[o;u]mⁿa- {KRDS, also some Hajong lects}. Diagnostic.

Reflexes of the *-m-* element are found in all KRDS lects and are unique to this area, as stated by Chatterji in the quote above. Having searched NIA data, I have found no such *-m-* element in third person plural pronouns anywhere else in NIA, with the exception of the most closely neighbouring Hajong lects (other Hajong lects further south are considerably different). These changes are unique and morphologically complex. For these reasons, [MI 31.] and [MI 32.] are diagnostic of a propagation event. They subgroup all of KRDS along with lects spoken by Hajong people in the neighbouring Garo hills (cf. §7.3.1).

5.6.5 Interrogative personal pronouns

Changes specific to the interrogative pronouns are localised to particular areas, and thus not of great significance for broader KRDS history. For TH the following two divergences from the proto Kamta system have been reconstructed:

[MI 33.] /ke/ 'INT.SG.NOM' +/-la/ 'PL' > /kela/ 'INT.PL.NOM' {TH: Muslims}. Nondiagnostic.

[MI 34.] /ka-/ 'INT.SG.OBL'+ /-ra/ 'PL.NOM' > /kara/ 'INT.PL.NOM' {TH: Hindus}. Nondiagnostic.

In BN, the suffix /-ba/ for indefinite pronouns is etymologically distinct from the broader KRDS suffix which is /ʌ; ɔ; o/. This divergence of BN away from the KRDS pattern once again brings it into closer conformity with Asamiya norms. The change is

morphologically specific, and thus diagnostic of contact relations with Asamiya. The indefinite affix /-ba/ used in BN and Asamiya is discussed by Kakati (1962:318):

> The affix -ba, *-bā* is often added to pronominal derivatives expressing manner or quality to suggest an indefinite sense; e.g. kɛnɛba, *kenebā*, konoba, *kono-bā*, zɛneba, *jene-bā*, kiba, *ki-bā*, etc. With -ba, the forms kono-, ka˜jo-, give an affirmative sense 'some body'.

[MI 35.] > /-ba/ 'INDF' in pronouns {BN, from SCA}. Diagnostic of contact relations with Asamiya.

Other divergences from the proto Kamta system are explained by phonological, or morphologically general changes, treated above in Chapter 4, or §5.6.1, respectively.

5.6.6 Relative personal pronouns

Among the relative pronouns, all that remains to be mentioned is a couple of localised analogical changes in RP:

[MI 36.] Analogical extension of /-mra, -ma/ to relative plural pronouns /zamra, zama-/ {RP}. Diagnostic.

The fact that the elements /-mra/ and /-ma/ have been innovatively extended in RP to relative plural function adds support to the analogical explanation given in §5.6.4 for the presence of the *-m-* element in third person plural pronouns across KRDS. This change (as for [MI 37.] below) is diagnostic of a PE based on ecological distinctiveness and linguistic complexity of the morphological conditioning.

A further analogical change again concerns the relative pronouns in RP:

[MI 37.] Extension of DIST/PROX distinction to the relative plural pronouns: /zamra/ 'REL.DIST.PL.NOM' versus /zemra/ 'REL.PROX.PL.NOM' {RP}. Diagnostic.

Together, [MI 36.] and [MI 37.] constitute a restructuring of the relative plural pronouns based on the model of the third person plural pronouns—distinguishing nominative versus oblique functions, singular versus plural number, and distal versus proximal location.

5.7 Adjectival and adverbial pronominal derivatives

In addition to the personal pronominals, KRDS has pronominal derivatives in both adjectival and adverbial categories. These forms also enter into paradigmatic relations, distinguishing proximal, distal, interrogative, relative, and sometimes anaphoric categories for each pronominal base. There are two systems of adjectival pronominals—quality and quantity—and multiple systems of adverbial pronominals including temporal, locational and directional pronominals.

The pronominals involve two morphemes: a deictic compounded with a nominal base. The deictics thereby recur across all pronominal systems, and some introductory comments on them are in order. The deictic forms for KRDS and some other Magadhan languages are displayed in Table 5.28.

Table 5.28: Magadhan deictic forms

	Proximal	Distal	Interrogative	Relative	Anaphoric
KS	i-, wɛ-	u-, wo-, ʌ-,	kɛ-, kə-	ʤɛ-	[a]
RL	hi-, ɛ(i)-	hu-, ʌ(i)-, sɛi-	ki-, kun-	ʤɛ(i)-	
MH	i-, ɛ(i)-	u-, ʌ(i)-	kɛ-, kun-	ʤɛ(i)-	
TH	ɛ-, ei-	ɔ-, oi-, ʃɛ-	kɛ-, kun-	ʧɛ-, ʧɔ-	
SH	ɛ-, ei-	ɔ-, oi-, ʃɛ-	kɛ-, kɔ-, ko-	zɛ-, zɔ-	
RP	ɛ-, ei-	ɔ-, oi-, ʃæ-, t̪ɔ-	kæ-, kɔ-, ko(n)-	ʤæ-	
BH	ɛ-, ei-	ɔ-, oi-, ʃɛ-	kɛ-, ki-, kɔ-, ko-, kun-	ʤɛ-, ʤi-, ʤɔ-	
BN	ɛ	o-, hɛ-, he-	kɛ-, ke-, ko(n)-	zɛ-, ze-	
Oriya	ɛ-	o-	kɛ-	ʧɛ-	ʃɛ-
SCB	æ-	o-, t̪æ-, t̪ɔ-	kɔ-, kæ-	ʧɔ-, ʧæ-	ʃæ-
SCA	ɛ-, ɔ-	tɔ-, tɛ-	kɔ-, kɛ-	zɔ-, zɛ-	xɛ-, xɔ-
Mth	e-, ə-	o-	kə-, ke-	ʧə-, ʧe-	t̪ə-, t̪e-
Bhj	(h)e-, (h)ə-	(h)o-	kə-, ke-	ʧə-, ʧe-	t̪ə-, t̪e-

[a] Anaphoric pronominals were not systematically collected as part of the KRDS data. Some KRDS lects always employ the DIST for ANP function, other KRDS lects have distinct DIST and ANP forms. Further data are required before these differences can be understood and historically explicated.

Proximal forms are marked by a front vowel, distal forms by a back vowel, interrogatives by an initial *k-, and relatives with an initial *ʧ-. All these features have been inherited into these lects as well as other NIA lects (cf. Chatterji 1926:829). The exact vowel quality in proximal and distal forms varies across KRDS, as well as in Magadhan languages more generally. The tendency for prothesis of a glide in western KRDS (KS, RL, MH)—for example /wə̃haj/ < *ɔ̃haj < *o̰haj ‘s/he’—is akin to the ‘Bihari’ lects, Maithili and Bhojpuri.

There is recurring variation in the vowel element of interrogative and relative deictics: for example SCB /kɔ-, kæ-/, and the cognate Maithili forms /kə-, ke-/. The wide distribution of this variation suggests a Magadhan inheritance. The back vowel variant *kɔ- is absent in Oriya as well as several KRDS lects. Given their non-contiguous locations, this is more likely the result of independent regularisation of inherited variation, rather than a propagation event.

The outcome of this short discussion is that the deictic forms inherited into Magadhan lects are not reconstructible to unique protoforms. Rather the inheritance includes variation within certain parameters: front vowels for proximal, back for distal, *kɔ- and *kɛ- for interrogative and *ʧɔ- and *ʧɛ- for relative forms.

The reconstruction of pronominal systems below focuses on differences in the *compounded noun* portion of the pronominals (e.g. /ei-**mon**/ 'this kind'), rather than on variation in the deictic element.

5.7.1 Adjectival pronominals of quality

The first set of adjectival pronominals are concerned with qualities of the referent, for example /**ɛnoŋ** nok/ 'this kind of man', /**kɛmun** asen/ 'how are you'. The nouns which are compounded with the deictics to create this pronominal system are as follows:

Table 5.29: Noun bases for pronominals of quality in KRDS, eMg, and Mg

pMg				*-sana*
p-eMg			*-manṭa*	*-sana*
pKmt	*-rɔkɔm	*-lak^{h}a	*-mɔṭɔn	
KS	-rʌŋ			
RL	-rʌŋ; -nʌŋ			
MH	-nʌŋ			
TH	-noŋ (H); rɔkom (M)		-mon (M)	
SH			-mɔṭɔn; -mɔn	
RP		-ŋka	-mon	
BH		-naxan	-mɔṭon; -mun	
BN		-ŋka		
Oriya			-mɔnṭɔ; mɔṭi	
SCB			-mon; -moṭ	-no
SCA				-ne
Mth				-hən
Bhj				-sən

Within KRDS there are three etymologically distinct bases used in pronominals of quality. These are reconstructed for proto Kamta as *-rɔkɔm, *-lak^{h}a, and *-mɔṭɔn (see Table 5.29) in accordance with the phonological correspondences in Chapter 4. Reflexes of the first etymon are found in the four westernmost lects. The reduction in form is not explained by regular phonological processes, and is a morphologically conditioned change:

[MI 38.] *-rɔkɔm > *-rɔŋ 'like, similar to' {KS, RL, MH, TH (Hindus, not Muslims)}. Diagnostic.

The morphological specificity of this change, geographical contiguity of range, and distinctiveness from surrounding lects all suggest this change to be diagnostic of a propagation event. The change of *r > n in TH, MH and variably in RL is not a regular phonological change (e.g. /rʌŋ/ 'colour'). However, there is a similarity between the nasalisation of this pronominal base (*rɔkɔm > *rɔŋ > *nɔŋ) and the nasalisation of the past tense marker *-il > /-in/ before a nasalised vowel (cf. §6.4.1.3). The nasalisation of

*l > n ([MI 68.]) before certain nasal features is much more geographically widespread than this change of *r > n. The nasalisation of *r in this pronominal extends the conditioning environment for medial nasalisation to cover the rhotic as well as the lateral. This extension is probably not diagnostic of a propagation event as the possibility of independent replication (given the pre-existing nasalisation process for *l) is quite high.

The second etymon, *-lak^{h}a, is likewise phonologically reduced by a morphologically-conditioned change:

[MI 39.] *-lak^{h}a > *-ŋka ‘like, similar to’ {RP, borrowed into BN}. Diagnostic of contact relations between RP and BN.

The nasalisation of *l is expected in RP by [PI 14.] because of the initial position of *l in *lak^{h}a as an independent noun. The presence of a nasal for *l in BN is phonologically irregular (see §4.3.11), and indicates that the lexeme is a loanword from RP into BN.

The third etymon *-mɔ̰ʈɔn is not unique to KRDS, but shared with modern Oriya and Bangla. This etymon was also present in early Asamiya as *-mata*, *mana* (Chatterji 1926:852, Kakati 1962:322) but it has been all but lost from the modern Asamiya language. The KRDS, early Asamiya and Bangla mix of inherited forms *-mɔnɔ, *-mɔ̰ʈɔ and *mɔ̰ʈɔn are alternative reflexes of a still earlier *mɔnʈ̰ɔ. The Oriya reflex /mɔnʈ̰ɔ/ is thus archaic.

[MI 40.] *mɔnʈ̰ɔ > *-mɔnɔ, *-mɔ̰ʈɔ ‘like, similar to’ {SCB, SCA, KRDS}. Probably diagnostic.

This change is old, attested in Bangla documents of the 14th century (SKK), as well as in the Asamiya writings of the late 15th century (authored by Sankara-Deva). Whether these forms have been lost in western KRDS and thus were part of a common Bangla-Asamiya-Kamta inheritance, or are instead to be accounted for by a more recent and limited propagation, must be decided on sociohistorical grounds in Chapter 7. Loss in western KRDS of the variation created by [MI 40.] would not be diagnostic of a propagation event.

5.7.2 Adjectival pronominals of quantity

The second set of adjectival pronominals is quantitative, for example /**ɛʈ̰ɛla** nok/ ‘this many people’, /**kɔʈ̰o**/ ‘how many’. These pronominals are more consistent across Magadhan lects than for the qualitative pronominals examined above.

Table 5.31: Comparison of pronominals of quantity in KRDS, eMg, and Mg

pMg	*-t̪e, -t̪a*	
p-eMg	*-t̪ɛ, -t̪ɔ	
pKmt	*-t̪ɛ, *-t̪ɔ	
KS	-t̪ə-la, -t̪ɛ-xan	
RL	-t̪(ɛ)-la, -t̪-kina	
MH	-t̪(ɛ)-la	
TH	-t̪o	
SH	-t̪o(-la)	
RP	-t̪o-ʈa, -t̪o-lʲæ, -knʲæ	
BH	-t̪o-la	
BN	-to-ɣila	
Oriya	-t̪ɛ	
SCB	-t̪o	
SCA	-te(-k), -tɒ-bor	-man
Mth	-t̪e-k	
Bhj	-t̪e-k	

With the exception of the Asamiya pronominals in /-man/, all these forms are cognate and constitute retentions. For discussion of the MIA and OIA etymology of affix *-t̪- see Chatterji (1926:855). Note that possible cognates of the KRDS protovariants *-(t̪ɛ, t̪ɔ) are found in early Maithili as -(*te˜, ta*).

It is not clear whether the /k/ element—pleonastic in Maithili, Asamiya and some of KRDS—forms part of the inherited pronominal material or is rather the result of independent replications of the same extension. The extension of quantitative pronominals with the various plural morphemes (e.g. /-la/, /-gila/) is noncomplex, and not diagnostic of a propagation event.

5.7.3 Temporal pronominals

The next few pronominal sets are adverbial rather than adjectival, and address temporal, locational or directional features. For the KRDS temporal adverbials all four forms—proximal, distal, interrogative and relative—are shown because of a change which affects some but not all of these forms. Cognates are listed in columns; due to the limited number of columns, noncognate forms are indicated by shaded cells.

Table 5.31: Comparison of temporal pronominals in KRDS, eMg, and Mg

	'now' PROX	'that time, then' DIST/ANP	'which time, when?' INT	'that time which, when' REL
p-eMg		*-k^hɔn		*-bɛ
pKmt	*ɛl^{ɳ}a	*ʃɛ-bɛ⌊a	*kɔun-bɛ⌊a	*ʤɛ(i)-bɛ⌊a
KS	al^{ɳ}a	u-xuna	kə̱t-k^huna	ʤɛ-xuna
RL	al^{ɳ}a	sɛi-βɛla	kun-bɛla; kun-k^huna	ʤɛi-βɛla; ʤɛi-k^huna; ʤʌb; ʤʌp
MH	al^{ɳ}a	ʌi-βɛla; u-βɛla	kun-bɛla	ʤɛi-βɛla; ʤʌb
TH	ɛla; ɛlan^{ɳ}e; ɛl^{ɳ}aj	ʃɛla; ʃɛlan^{ɳ}e	kun-bɛla	ʤebɛla; ʤeβɛla; ʤɛla
SH	ɛla	ʃɛla	kɔ̱tɔ-kkon	zɛla
RP	æla	ʃæla ṯɔ-kun	kon-bæla; kon-b^{ɳ}æla; kumbæla	ʤæla; ʤɔ̱t-k^hon; ʤebæla
BH	ɛla	ʃɛla	kun-bɛla	ʤɛla
BN	ɛla	hɛla	kon-bela; kun-bela	zela; sela
Oriya	-ṯɛ-bɛ⌊ɛ	-ṯɛ-bɛ⌊ɛ	-ṯɛ-bɛ⌊ɛ	-ṯɛ-bɛ⌊ɛ; -bɛ
SCB		-k^hon	-k^hon	-k^hon; -be
SCA			-ṯ^{h}ɒni; -hani	*-be*; *-we*[17]
Mth		-k^hən		-be
Bhj	-bera˜	-bera˜	-bera˜	-ʤun; -bera˜

[a] Anaphoric pronominals were not systematically collected as part of the KRDS data. Some KRDS lects always employ the DIST for ANP function, other KRDS lects have distinct DIST and ANP forms. Further data are required before these differences can be understood and historically explicated.

The KRDS temporal pronominals are consistently derived from the protonoun *bɛ⌊a 'sun, time' from OIA *vélā*. The proximal form 'now, this time' is also derived from *bɛ⌊a at some pre-proto Kamta stage. At the time of proto Kamta, however, it is most likely that the proximal form had become *ɛlha. Phonologically this reconstruction accounts for the distribution of /l^{ɳ}/ and /l/ in KRDS, with *l^{ɳ} > /l/ by [PI 9.]. In MH, RL and KS the contemporary form is /al^{ɳ}a/ 'now', the initial /a/ of which comes irregularly from *ɛ 'this', plausibly by *ɛ > *[ə] > /a/. This alteration is morphologically specific, and diagnoses a PE involving KS, RL and MH.

[MI 41.] *ɛl^{ɳ}a > /al^{ɳ}a/ 'now' {KS, RL, MH}. Diagnostic.

[17] Italics indicate a written form which has been romanised.

In other KRDS lects, there is a reduction of *bɛɭa in the anaphoric and relative functions:

[MI 42.] *bɛ > Ø in ANP and REL temporal pronominals {TH, SH, RP, BH, BN}. Diagnostic.

This change accounts for forms such as /ʃɛla/ < *ʃɛbɛla 'then'. This change does not require a complex series of changes as in the case of [MI 41.]; nevertheless the conditioning has a degree of complexity (ANP and REL, but not INT) which is uniform across a contiguous area, justifying the reconstruction of a propagation event. Variation in the relative forms in TH and RP need not affect the formalisation of this change. The presence of the fuller form /ʤebɛla/ alongside the reduced form /ʤɛla/ is probably due to the re-creation of the fuller form by analogy with the interrogative form /kun-bɛla/ 'when?'.

Cognate pronominals are found in Bhojpuri /-bera˜/ and Oriya /-bɛɭɛ/ *beḷe*. The Bhojpuri substitution of /r/ for *l, is consistent with its Magadhan inheritance (cf. Masica 1992:186).

The pronominal element *-bɛla < OIA *velā* is not cognate with the pronominal element /-be/ < OIA *-va* found in pronominals across the Magadhan languages. Examples of the latter cognate set include the early Asamiya forms given in Table 5.31 as *e-be* etc. (after Chatterji) and *ewe* etc. (after Kakati).

Temporal pronominals derived from the reflex of OIA *kṣana* > /-(k)k^{h}ɔn/ are found in all Magadhan languages according to Chatterji (1926:857). The use of nominal base /-k^{h}una/ in KS and RL may be Maithili influence, but could also constitute retentions. The RP form /ʈɔkun/ is likely to be a Bangla loanword, and SH /kɔʈɔkk^{h}on/ possibly a Sanskritism (Tatsama/semi-Tatsama). However, in both cases the possibility of retention from MIA must first be ruled out—a task which awaits further study.

5.7.4 Locational pronominals

The second set of adverbial pronominals refer to the location of an event. This set is also reasonably uniform across KRDS.

Table 5.32: Comparison of locational pronominals in KRDS, eMg and Mg

pMg	*-*ɦā˜i*				
p-eMg	*-*ɦā˜i*				
pKmt	*-ʈ^{h}ɛ	*-ʈ^{h}ɛ-kuna			
KS		-ʈ^{h}in; -ʈ^{h}ina			
RL		-ʈ^{h}a; -ʈ^{h}ina			
MH	-ʈ^{h}ɛ	-ʈ^{h}ina			
TH	-ʈ^{h}e	-ʈ^{h}e-kona			
SH	-ʈ^{h}e				
RP	-ʈe	-ʈe-kona			
BH	-ʈe	-ʈi-xun^{j}æ			
BN	-tɛ				
Oriya	-ʈ^{h}i				
SCB			-k^{h}ane < CLF		
SCA				-t < LOC	
Mth				-tə(e) < LOC	
Bhj	-ʈ^{h}ən, -ʈ^{h}in				-ha˜

The Asamiya locational pronominals are based on the locative case ending /-t/ < *-ɔṭɔ, and are not cognate with the KRDS pronominals. The same goes for Maithili /-tə/, which Jha (1958:§555)derives from OIA *-tra*.

The locational pronominals in KRDS are reconstructed as derivatives of *-ʈʰɛ, in turn cognate with Oriya /-ʈʰi/, as well as Northern and Western Bhojpuri /-ʈʰən, ʈʰen, ʈʰin/. Middle Bangla of the Sri Krishna Kirttana (14th century) has ঠাই *-ṭhāi*. The most probable form for the proto-eastern Magadhan stage is reconstructed as **-ṭhā̃i* (following Chatterji 1926:769) < *sthā́man* (cf. Turner 1962–66: id. 13760) from which we derive Oriya and KRDS forms by changes whose regularity has not been tested:

[MI 43.] *-ṭhā̃i > /-ʈʰi/ ‘place’ {Oriya}. Diagnostic value unknown.

[MI 44.] *-ṭhā̃i > *-ʈʰɛ ‘place’ {KRDS}. Diagnostic value unknown.

Alongside KRDS cognates of *-ʈʰɛ (<*-ṭhā̃i) there are extended forms, with /-ina/ suffixed in the west and /-kuna, kona/ in the centre and east, which require some discussion.

There are two distinct etymologies possible for /-ina/ suffix found in MH, RL and KS. Firstly, it may be cognate with Bhojpuri /-ʈʰən, -ʈʰin/, and constitute an inheritance alongside *-ṭhā̃i from the common Magadhan period. Alternatively, it may be cognate with the suffix /-kuna, kona/ found in central and eastern KRDS lects TH, RP and BH.

Tiwari (1960:150) reconstructs the etymology of the Bhojpuri forms as follows:

> The origin of **-ṭhan, -ṭhen, -ṭhin** and **-ṭhẽ** forms ... is possibly the pronominal base √*sthā* + the locative affix **hĩ** or **ahĩ**. These forms can be compared with the dialectical Bengālī forms *sēṭhi, ēṭhi, jēṭhi,* and with *ṭhi-* forms in Oṛiyā.

Based on Tiwari’s reconstruction, the /n/ element in Northern and Western Bhojpuri is cognate with the nasal element of **-ṭhā̃i*. This would constitute a highly irregular and clumsy etymology for the /-ina/ ending in MH, RL and KS.

A simpler and neater etymological explanation for the /-ina/ is through considering possible cognacy with /-kuna, kona/ < *-kuna in TH, RP and BH. The simplicity of this explanation is that /-ina/ and /-kuna, kona/ are given for contrasting KRDS lects. The process would be as follows:

*-ʈʰɛ-kuna ‘place’ > *-ʈʰikuna (by regressive vowel raising, [PI 20.])

> *-ʈʰikna (by changes to medial high vowels, see §4.4.6) > /-ʈʰina/ ‘place’.

The only parts of this process that are not accounted for by phonological changes are the initial compounding of *-ʈʰɛ-kuna, and the final reduction > /-ʈʰina/.

[MI 45.] *-ʈʰɛ + kuna > *-ʈʰɛkuna ‘place’ as a base of locational pronominals {KRDS}. Diagnostic.

[MI 46.] *-ʈʰikna > /-ʈʰina/ ‘place’ {MH, RL, KS} (after [PI 20.] and [PI 30.]–[PI 33.]). Diagnostic.

The first change is, to my knowledge, unique to KRDS. Based on this uniqueness, coupled with the morphological specificity of the change, it is diagnostic of a propagation event.

The reduction of *kn to /n/ is not a phonologically regular change in these lects (cf. MH /nukni/ ‘louse’), but specific to this morpheme in this pronominal set. The specificity increases the complexity of the change, which is uniform across a contiguous area, and diagnostic of a propagation event.

This concludes the reconstruction of changes in the KRDS pronominal sets, and changes in nominal morphology more generally.

5.8 Summary of diagnostic innovations in nominal morphology

The following morphological changes have been reconstructed in this chapter to be either (i) diagnostic of propagation events; (ii) supportive of other diagnostic changes; or (iii) of unclear diagnostic value to be evaluated further in Chapter 7.

[MI 1.] > /-d̪-/ ‘PL.OBL.AN’ {SCB} (before AD 1500). Diagnostic.

[MI 2.] *pronoun-GEN(-a)* + *noun of multitude* ‘plural pronoun’ > pronoun-GEN(-a) ‘plural pronoun’ {middle Bangla, early Asamiya, KRDS}. Supportive, not diagnostic.

[MI 3.] /-[e]ra/ ‘PL.NOM’ in pronouns > /-[e]ra/ ‘PL.NOM.AN’ in general nominal morphology {SCB} (by the C15th). Diagnostic.

[MI 4.] > /-[ɔ]r/ ‘GEN’ {BN, from SCA}. Supportive of contact relations with Asamiya.

[MI 6.] *-ɔkɔ ‘DAT’ + *-ɛ ‘INS-LOC’ > /-ke/ ‘DAT’ {SCB, …}. Supportive, not diagnostic.

[MI 7.] *-[ɔ]t̪ ‘LOC’ + *-ɛ ‘LOC-INS’ > /-t̪e/ ‘LOC’ {SCB, Lodha} (before AD 1400). Probably diagnostic.

[MI 8.] > /sɛ/ ‘INS’ {RL, KS from Hindi/Bihari}. Diagnostic of contact relations of diglossia with Hindi.

[MI 9.] > *t̪hakia ‘ABL’ {SCB, TH, SH, RP, BH}. Tentatively diagnostic of contact relations with SCB through diglossia.

[MI 10.] > /sɛ/ ‘ABL, CMP’ {RL, KS, MH}. Diagnostic of contact relations through diglossia with Hindi.

[MI 11.] > /pɔra/ ‘ABL’ {BN, SCA}. Diagnostic of contact relations with Asamiya.

[MI 12.] > /kɔi/ ‘CMP’ {BN, SCA}. Diagnostic of contact relations with Asamiya.

[MI 21.] Pronouns with function PL.OBL are extended to general plural function (thus including PL.NOM) {RL, MH}. Diagnostic.

[MI 22.] > *ʃɔmbɔ ‘PL.OBL’ in pronoun declension {KRDS, early Asamiya} (tentatively p-Kamrupa stage). Supportive, not diagnostic.

[MI 23.] *hama-[ʃɔmba-] ‘PL.OBL’ > *hama-[ʃɔma-] > *hama-[ʃa-] ‘1PL.OBL’, and the equivalent changes in other pronoun declension, e.g. *t̪omɲa-[ʃa-] ‘2PL.OBL’, etc. {KRDS, ?early Asamiya}. Diagnostic value unknown.

[MI 25.] /mɔj, tɔj/ ‘I, you’ {BN}. Diagnostic of contact relations with Asamiya.

[MI 28.] > /ami/ ‘we’ {BN}. Supportive, not diagnostic, of contact relations with Asamiya.

[MI 30.] > /apuni/ ‘2H.NOM’, /apona-/ ‘2H.OBL’ {BN}. Supportive, not diagnostic, of contact relations with Asamiya.

[MI 31.] *mɲra reinterpreted as ‘PL.NOM’ in pronoun system, and extended as such to third person *[o;u]mra {KRDS; also some Hajong lects}. Diagnostic.

[MI 32.] *mɲa- reinterpreted as ‘PL.OBL’ in pronoun system, and extended to third person *[o;u]mɲa- {KRDS; also some Hajong lects}. Diagnostic.

[MI 35.] > /-ba/ ‘INDF’ in pronouns {BN, from SCA}. Diagnostic of contact relations with Asamiya.

[MI 36.] Analogical extension of /-mra, -ma/ to relative plural pronouns /zamra, zama-/ {RP}. Diagnostic.

[MI 37.] Extension of DIST/PROX distinction to the relative plural pronouns: /zamra/ ‘REL.DIST.PL.NOM’ versus /zemra/ ‘REL.PROX.PL.NOM’ {RP}. Diagnostic.

[MI 38.] *-rɔkɔm > *-rɔŋ ‘like, similar to’ {KS, RL, MH, TH (Hindus, not Muslims)}. Diagnostic.

[MI 39.] *-lak^{h}a > *-ŋka ‘like, similar to’ {RP, borrowed into BN}. Diagnostic of contact relations between RP and BN.

[MI 40.] *mɔnt̪ɔ > *mɔnɔ, *mɔt̪ɔ ‘like, similar to’ {SCB, SCA, KRDS}. Probably diagnostic.

[MI 41.] *ɛl^{ɲ}a > /al^{ɲ}a/ ‘now’ {KS, RL, MH}. Diagnostic.

[MI 42.] *bɛ > Ø in ANP and REL temporal pronominals {TH, SH, RP, BH, BN}. Diagnostic.

[MI 43.] *-ṭhā̃i > /-ʈ^{h}i/ ‘place’ {Oriya}. Diagnostic value unknown.

[MI 44.] *-ṭhā̃i > *-ʈ^{h}ɛ ‘place’ {KRDS}. Diagnostic value unknown.

[MI 45.] *-ʈ^{h}ɛ + kuna > *-ʈ^{h}ɛkuna ‘place’ as a base of locational pronominals. {KRDS}. Diagnostic.

[MI 46.] *-ʈ^{h}ikna > /-ʈ^{h}ina/ ‘place’ {MH, RL, KS} (after [PI 20.] and [PI 30.]–[PI 33.]). Diagnostic.

The sociohistorical conditioning of propagation of these changes is examined in Chapter 7.

6 *Reconstruction of verbal morphology*

6.1 Introduction

This chapter is a reconstruction of changes in two key areas of KRDS verbal morphology: tense/aspect, and personal endings. The reconstruction of verbal morphology would be made more complete by a thorough study of negation and participial morphology, but these tasks require further data than are presently available. The verbal forms described and compared herein are all finite in function, with the exception of the perfective and the infinitive. These two nonfinite forms are included because they figure in the finite verbal constructions of some lects (e.g. the continuous aspect construction).

Finite verb constructions in KRDS may be simple or compound. Simple verbs consist of a verb stem suffixed with tense/aspect morphology and agreement (AGR) endings: Verb–Tense/Aspect–AGR.

In almost all KRDS lects, the finite verb agrees with the grammatical subject with respect to person and number. The exception to this rule is central Jhapa Rajbanshi, which has the standard subject agreement, but also a system of *secondary or patient agreement* between the verb and the patient of the clause. This unique (for KRDS) morphological system is carefully described in Wilde (2008:147–172).

Compound verbs consist of a (semantically) main verb (Verb1) as participle, followed by a simple and finite auxiliary verb (Verb2): Verb1–Participle + Verb2–Tense/Aspect–AGR.

The set of auxiliary verbs is limited,[1] and the auxiliary meaning is different to the independent verbal meaning of the lexeme. The function of auxiliary verbs is stated by Masica (1991:326):

> Partially emptied of their lexical content, these modify the meaning of the main verb in various ways not unrelated to that content, which might best be described as *manner-specification*. (emphasis original)

Masica designates the usual role of these auxiliary verbs in compound formations as *Aktionsart*—'they belong more to the domain of derivation, that is, to lexicon, than to grammar' (ibid.:268, 326ff.; cf. Goswami and Tamuli 2003:429ff.). Given the NIA generality of this phenomenon, we can expect that auxiliary verbs are used for derivational

[1] In Indo-Aryan studies, auxiliary verbs may be alternatively termed: intensifiers, operators, explicators or vectors.

Aktionsart categories beyond the functions described in this chapter. One auxiliary verb receives special attention in this chapter (see §6.2.4) because it has been grammaticalised for continuous aspect in RP, BH and BN.

Tense/aspect morphology is described and reconstructed in §6.2. Following this, the agreement systems of the eight KRDS lects are described in §6.3, and historical changes are reconstructed in §6.4. It is preferable to treat the verbal morphology in this order because an understanding of Tense/Aspect history informs the reconstruction of agreement history.

6.2 Tense-Aspect morphology: description and reconstruction

In a comparative study of under-described lects, such as this one, it is not possible to do full justice to describing the functional relations within the tense/aspect system of each lect. Of the lects examined here RL, MH and BH have been subject to modern linguistic description—though with the exception of Wilde (2008) the descriptions have not been of the standard of a reference grammar. We may expect that further and more thorough grammatical descriptions of KRDS lects will be available in the near future. Such studies may be used to test the diachronic arguments outlined here. In the meantime, comparison is made of tense/aspect formations with broadly similar (if not identical) functions. It is possible to reconstruct formal changes in the systems, as well as some broad functional changes, even while we await more detailed descriptions of verbal semantics and functions in much of KRDS.

The verbal formations that are reconstructed below for proto Kamta are given in advance in Table 6.1 to aid the reader in following the arguments that follow. The use of 'proto Kamta' to denote this stage is justified on historical grounds in §7.3.1. Different sets of agreement endings pertain in different verbal formations, hence AGR.I, AGR.IIA, etc. The synchronic use of multiple sets of agreement endings is explained in §6.3, and the historical origins explained in §6.4.

Table 6.1: Verbal formations reconstructed for pKmt

	Indefinite aspect	Perfective aspect	Continuous aspect
Past	*VERB–il–AGR.IIA	*VERB–i tʃʰil–AGR.IIC	
Present	*VERB–AGR.I	*VERB–i–tʃʰ–AGR.I	*VERB–ɛ–tʃʰ–AGR.I
Future	*VERB–i[b,m][a]–AGR.IIB		

[a] The allomorphy is explained historically in §6.2.6.

With the exception of the 'present perfective' data, the tense/aspect formations were tested using model texts collected at all eight sites. The data for 'present perfective' were collected through more controlled and leading elicitation, and therefore may not be as reliable an indication of language use as the data for the other categories. There seems to be an overlap in function between the 'present perfective' and the 'simple past' and 'past perfective' in §6.2.5. They are partially interchangeable in certain discourse contexts, but the contexts permitting such interchange are yet to be studied. In fact, Wilde finds tense marking to be neutralised in narrative, with the tense markers taking on pragmatic functions rather than temporal reference (2008:Ch.8). This finding is yet to be tested for

other KRDS lects. In the absence of discourse studies for all the KRDS lects, the labels applied to tense and aspect categories should be considered approximations of non-narratival usage.

Before beginning the description and reconstruction of finite verbal formations, a few pieces of derivational morphology that figure frequently in that discussion require closer analysis.

6.2.1 Perfective morphemes

In KRDS there are three kinds of perfective morphemes. These morphemes are structurally and formally distinct in at least some KRDS lects. The forms are shown in Table 6.2, with column headings explained below.

Table 6.2: Perfective morphemes across KRDS and reconstructed for pKmt

	PFV in adverbial clauses	PFV in compound verbs [a]	PFV in simple verb formations
pKmt	*-ia	*-ia	*-i
KS	[b]	-(i)ɛ	-i
RL	-[j]ɛ=k^{h}una	-(i)ɛ	-i
MH	-[j]ɛ=nɛ	-(i)ɛ	-i
TH	-hene	-e →	-[i;e]
SH	-ia; -iɛ	-i ←	-i
RP	-iæ	-i ←	-Ø
BH	-ia; -iɛ	-i ←	-Ø
BN	-ia	-ia	-i

[a] In his study of central Jhapa Rajbanshi, Wilde divides the middle category of compound verbs into two by differentiating 'quasi-aspectual compound verbs' from 'derivational compound verbs' (see Wilde 2008:211ff.; and especially 229, fn.70). The former are semi-grammaticalised, involving one of only four quasi-aspectual auxiliary verbs. These four auxiliary verbs do not figure derivational compounds, which are idiomatic and belong to the domain of lexicon rather than grammar. This semantic and syntactic distinction should be further explored for the other KRDS lects.

[b] Data missing.

The most grammaticalised of these three perfective categories is in the rightmost column. This morpheme occurs in present perfect and past perfect verbal formations directly after the verb stem and followed by the tense morphology (cf. §6.2.4 and §6.2.5 respectively). For example (from MH), /dɛk^{h}-**i**-ʧ-u/ 'see-**PFV**-PRS-1SG'='I have seen'. This morpheme is /-i/ across KRDS, except in TH, and it has been lost in RP and BH. The corresponding TH morpheme is /-e/, with the allomorph /-i/ resulting from regressive vowel raising when a high vowel occurs in the following syllable. This difference in form between TH and the rest of KRDS is not explained by phonological changes, but is the outcome of morphological changes that remodel TH's verbal morphology in favour of SCB norms (see further §6.2.4). Based on the broad distribution of /-i/ across KRDS, *-i is reconstructed as the proto Kamta form with 'perfective' function in this structural position. This reconstructed morpheme can be seen in Table 6.1 as part of the perfective formations.

The second category of perfective markers attaches to *main verbs in compound verb formations* (see description above for the structure of these formations). For example, in Bhatibari: /mui ḍ̥ek^{h}**-i** p^{h}ɛla-s-uŋ/ ‘I have seen (it)’. In this example, the main verb is /ḍ̥ɛk^{h}-/ ‘see’, suffixed with the perfective marker /-i/; the auxiliary verb is /p^{h}ɛla-/ ‘throw’, conjugated for tense/aspect and agreement. Across KRDS, there is greater variation of perfective marking in this compounded position (cf. the middle column of Table 6.2). This variation is not explained by the phonological correspondences constructed in Chapter 4. Two hypotheses are possible. Firstly, western KRDS and eastern KRDS forms (which are also cognate with the adverbial perfective markers) could be irregular reflexes of *i in this position that result from a morphologically-specific sound change. However, this leaves unexplained why such a change did not apply in the case of the perfective in simple verbs (rightmost column of Table 6.2). Moreover, the lowering of *i to /(j)ɛ/ in the western lects is not well motivated phonologically.

A better option for the etymology of /(j)ɛ/ in compound verbs is suggested by comparing it with the adverbial perfective forms in the leftmost column. The eastern and western KRDS lects use the same perfective marker in both compound verbs and adverbial clauses. The central KRDS lects on the other hand use the inherited perfective *-i from the simple verbs in compound formations, and /-ia; -iɛ; -iæ/ < *-ia in adverbial clauses. The most economical diachronic explanation of this divergence is that the inherited perfective marker in *compound verbs* was *-ia (the same as in adverbial clauses), and that this morpheme was replaced in the central KRDS lects by the simple verb perfective *-i. In RP and BH the inherited *-i was then lost in simple verbs.

[MI 47.] *-i ‘PFV’ in simple verbs > /-i/ ‘PFV’ in both simple and compound verb constructions {BH, RP, SH} (before [MI 56.]). Diagnostic.

This change bears partial similarity with Asamiya, which has regularised /i/ as perfective in all three of the distinct structural positions outlined for KRDS in Table 6.2 (simple verbs, compound verbs and adverbial clauses). The Asamiya change is a different change to [MI 47.] which is more tightly constrained and excludes perfective marking of adverbial clauses. On the basis that the morphological conditioning of [MI 47.] is complex, it is diagnostic of a propagation event.

The third structural position of perfective morphemes is attached to *nonfinite verbs in adverbial clauses*. This position is common across Indo-Aryan languages and the morpheme is termed the ‘perfective conjunctive’ or ‘conjunctive participle’ (cf. Masica 1991:323). In some lects this morpheme can indicate instrumental and causative rather than simply perfective functions (Wilde 2008:230–231).

The reconstructed perfective marker in this adverbial position is *-ia. Note that the corresponding morphemes are not all regular reflexes of *-ia. Nevertheless there is enough similarity across the attested forms to make cognacy fairly sure, and to justify the hypothesis of a morphologically-specific sound change. Raising of the final vowel of the suffix *-ia occurs in the central KRDS lects (> /iæ; iɛ/ ‘PFV’); in central KRDS, this is a phonologically regular change (cf. §4.4.2). However, the raised western reflexes cannot be similarly explained because progressive vowel raising is not a phonological process in those lects. Instead the raising in these lects seems to be connected to the re-phonemicisation of the *i element of *-ia as a glide: *-ia > *[-ja, jɛ] > /-(j)ɛ/. The brackets in the final form indicate that the glide element is variable. The change of *a > /ɛ/ in this

environment is not attested by multiple correspondences, but nonetheless it is a plausible sound change. It is more plausibly motivated than the alternative etymology of *i > /(j)ɛ/ which was rejected above. The following change is reconstructed for the western lects as a morphologically-specific change:

[MI 48.] *-ia ‘PFV’ > *[-ja; jɛ] > /-(j)ɛ/ ‘PFV’ {KS, RL, MH, TH}. Diagnostic value unclear.

A similar change has affected the Bangla inherited perfective: *-ia > /-e/ ‘PFV’. Bangla influence in this respect is sociohistorically plausible in the case of TH which is within the modern Bengal sociopolitical zone and has undergone other changes in common with Bangla (e.g. prosodic vowel raising, reconstructed to be a post-1800 change in §7.5). However, the lects KS, RL and MH are outside the Bengal zone and tend to be influenced by diglossia with Hindi, not Bangla. Therefore at least for KS, RL and MH this change seems to be unrelated to the structurally similar change in Bangla. The case of TH is ambiguous because the change could have been a common propagation with KS, RL and MH during their common period of development (AD 1550–1787, cf. §7.4.2), or alternatively could have been a more recent innovation (post-1800) due to diglossia in Bangla. This ambiguity cannot be resolved on linguistic criteria, and the diagnostic value of [MI 48.] is listed as unclear.

All three positions of perfective marking are illustrated by the following example from MH: /mui ḍɛkʰ-**ɛnɛ** gʱur-**jɛ** as-**i**-ʧ-u/ ‘Having seen (it), I came back’. Firstly, the perfective is suffixed to the simple verb /as-/ ‘come’; in this position the perfective is /-i/. Secondly, the perfective is attached to the verb /gʱur-/ ‘turn’ in the compound construction. In MH, the perfective is /-(j)ɛ/ in this position. Thirdly, a perfective marker /-(j)ɛnɛ/ is also attached to the verb root /ḍɛkʰ-/ ‘see’. The adverbial clause /ḍɛkʰ-**ɛnɛ**/ ‘having seen’ is a subordinate clause to the main clause ‘I came back’. The semantic relation between the adverbial clause and main clause is of ‘a *succession of actions or events* done by or with reference to the *same* subject’ (Chatterji 1926:1003). This adverbial relation between clauses differs from that of a compound verbal construction which describes a single event.

These perfective markers are inherited, with cognates found in other Magadhan languages (and perhaps further afield in NIA also):

> The conjunctive -i is derived from M.I.A. *-ia* < *O.I.A. -ya.* In [Bangla] it appears in the strengthened form *-iyā.* In [Early Asamiya] both the forms in *-i, -iyā* are found. (Kakati 1962:365)[2]

The phonological reflex of MIA *-ia* is proto Kamta *-iɔ, which becomes /-i/ by loss of final *ɔ. This MIA suffix *-ia* was extended by *-a to give early Asamiya and high literary Bangla *-iā*, with SCB /-e/ derived from that extended suffix. The extension of *-ia* > *-iɔ + *-a > *-ia parallels the extension of many proto Kamta nouns that end in *ɔ with the nominal suffix *-a (cf. §4.4.11). That nominal suffixing process is of unclear diagnostic value because it is so broadly distributed across NIA. Any hypothesis regarding the diagnostic value of extending the perfective marker with *-a should be based on a consideration of the diagnostic value of suffixing nouns ending in *ɔ with *-a. Consequently the following change is currently listed as ‘diagnostic value unclear’. This change also occurs in early Maithili (Jha 1985 [1958]).

---

[2] Note that the ‘*a*’ in the MIA form *-ia* is a short ‘a’, thus corresponding with KRDS /ɔ/ not long ‘ā’ which corresponds with KRDS /a/.

[MI 49.] *-iɔ 'PFV' + -a 'nominal suffix'> *-ia 'PFV' {Middle Bangla, Early Asamiya, early Maithili, KRDS}. Diagnostic value unclear.

This affix is historically connected to the contemporary Bangla morphophonemic process which dictates 'high' and 'low' variants of verb stems in different morphological environments. For example /aʃ-/ 'come' is the low alternate, and /eʃ-/ 'come' the high alternate. The high alternate occurs when suffixed by the perfective participle /-e/ < *-ia. This distribution led Chatterji to propose a transposed *i from *-ia (by [PI 30.]) to be the historical cause of 'strong' raising of the preceding vowel. (It is termed 'strong' because it affected all vowels including *a, which is not the case for the general regressive harmonic process, cf. §4.4.1) Diachronically, the 'high' alternate of the verb stem has been retained before the perfective participle, even though the phonological trigger has been lost by the change *-ia > /-e/. The synchronic result is a Bangla morphophonemic process which is absent from KRDS and Asamiya.

6.2.2 The Infinitive

Infinitive forms of verbs are found in KRDS as verbal complements (e.g. /mui **ʤa-ba** ʧahaʧu/ 'I want **to go**' {MH}) as well as in broader nominal uses suffixed by genitive case (e.g. /**d̪ɛkʰ-iba-r** paʧʰɔt̪ ʤam/ 'after **seeing** (it) I will go' {MH}). The forms for the infinitive across KRDS are shown in Table 6.3.

Table 6.3: Infinitive morphemes across KRDS and reconstructed for pKmt

	Infinitive morpheme
pKmt	*-iba
KS	-na
RL	-ba
MH	-ba
TH	-ba
SH	-bar
RP	-bær
BH	-bær, -ir
BN	-ba

The forms are partially cognate across KRDS, with the exception of KS which is cognate instead with the Hindi infinitive, and constitutes a loan. This change is diagnostic of KS's relation to Hindi through diglossia.

[MI 50.] *-iba 'INF' replaced with /-na/ 'INF'. Diagnostic of contact relations through diglossia with Hindi.

Otherwise, the infinitive participle is /-ba/ across the eight KRDS lects, except in SH, RP and BH where it is /-bar, -ir/. In these three lects a nominalised form in genitive case has been reanalysed as infinitive. This is a diagnostic change for these central KRDS lects.

[MI 51.] *-iba 'INF' + *-[ɛ]r 'GEN' > *-ibar 'INF'{SH, RP, BH}. Diagnostic.

The allomorph /-ir/ only occurs in BH and the surrounding areas of Cooch Behar and Dhubri. In this lect /-ir/ attaches to verbs ending in a consonant, for example /dɛkʰ-ir/ 'to see', and /-bar/ attaches to verbs ending in a vowel, for example /ʤa-bar/ 'to go'. This allomorphy is distinct to BH (and the adjacent areas just mentioned); it is morphologically and phonologically specific, and considered a diagnostic change.

[MI 52.] *-ibar 'INF' > /-ir/ / C_ {BH}. Diagnostic.

The proto Kamta infinitival suffix *-iba is cognate with /-iba/ in Oriya and /-ibɒ/ in Asamiya, and thus a pre-proto Kamta inheritance. The *i element is regularly or variably lost in all KRDS lects, probably in association with the changes described in §4.4.6. The Bangla infinitival suffix /-[i]ṭe/ is not cognate, and constitutes an innovation.

6.2.3 Imperative (and subjunctive?) and present indefinite

The imperative and present indefinite conjugations are treated together in this section by virtue of their morphosyntactic similarities, namely: (1) they are finite conjugations (i.e. they occur in finite clauses); but (2) they lack overt tense and aspect marking. In these formations the agreement suffixes attach directly to the verb stem, as in /kɔr-**iʃ**/, 'do-**2SG**'='you(SG) do' (present indefinite, from BH).

With no overt tense/aspect morphology to differentiate imperative from present indefinite, and with detailed semantic analysis of verbal conjugations beyond the scope of the present study, it is difficult at present to properly differentiate these two conjugations. Furthermore, it may be necessary to distinguish subjunctive from both imperative and present indefinite (as per Wilde [2008:148ff.]). Thus what is here termed 'first person imperative' (e.g. /mui kɔr-oŋ/ 'I do' or '(let) me do') may alternatively be 'first person subjunctive'. Moreover, the use of the 'second person indefinite' at times suggests rather a subjunctive (e.g. /kɔr-iʃ/ '(let) you do'. It remains unclear then, whether the verbal distinctions involved are two-way (imperative/indefinite) or three-way (imperative/subjunctive/indefinite). This facet of the KRDS verbal morphology is under-described in the present study and requires further study. The reliability of the reconstruction, however, is not much affected because the reconstruction focusses on morphological forms rather than precise verbal semantics and clausal functions.

In the rest of this section and chapter, second and third person endings are classed as either imperative (AGR.IMP) or present indefinite (AGR.I). This division may turn out to be unnecessary with regards the first person, as there is only one form; on the face of it, however, distinct imperative (or subjunctive) and indefinite functions seem to be present even for the first person. (Note that Masica [1991] prefers 'General Unspecified' and 'Present Habitual' to 'Present Indefinite'.)

In his study of central Jhapa Rajbanshi (most similar to RL in this study), Wilde (2008) finds the construction V-AGR.I to denote *past habitual* rather than *present indefinite* as it is labelled here. The temporal reference of this conjugation in narrative discourse is complicated, and in some contexts—for some lects—may indeed favour a past tense reading. However, my general study of KRDS lects thus far inclines me to denote the basic sense as present rather than past reference—with past tense a secondary, and discourse

dependent sense. This analysis is in accord with Bangla and Asamiya studies more generally. The aspectual difference between Wilde's 'habitual' and my 'indefinite' (or 'unspecified') is a slim one, and goes to a level of semantic exactitude beyond the scope of this study. Wilde's discussion of verbal semantics in central Jhapa Rajbanshi is more detailed than here, but—as he acknowledges—not conclusive on all matters.

The agreement suffixes which attach to the imperative and present indefinite are listed for each of the eight sample KRDS lects in §6.3.

Table 6.4: Imperative and present indefinite formations in KRDS and pKmt

	Imperative	Present indefinite
pKmt	*-AGR.IMP	*-AGR.I
KS	-AGR.IMP	-AGR.I
RL	-AGR.IMP	-AGR.I
MH	-AGR.IMP	-AGR.I
TH	-AGR.IMP	-AGR.I
RP	-AGR.IMP	-AGR.I
SH	-AGR.IMP	-AGR.I
BH	-AGR.IMP	-AGR.IA
BN	-AGR.IMP	-AGR.I

The structure of imperative and present indefinite formations is uniform across KRDS, as well as Asamiya, Bangla and Oriya. The agreement endings used in these conjugations are also cognate beyond KRDS in other eastern Magadhan lects (see §6.4). With cognate suffixes in identical structural positions, it is most likely that both these formations are inherited from proto Magadhan into proto Kamta. These imperative and present indefinite formations may be traced further back in history to the imperative mood and the present indefinite of OIA (cf. Chatterji 1926:864). The structural dissimilarity in BH, which has the present indefinite conjugated with a partially distinct 'AGR.IA' system (rather than AGR.I) is explained diachronically in §6.4.

6.2.4 Present continuous and present perfective

There are two 'present definite' tense formations, which along with the 'present indefinite' dealt with above, complete the present tense conjugations for KRDS. The present perfective conjugation is used in KRDS for completed events where the completion is relevant to, or temporally proximate to, the (discourse-defined) present moment. The present continuous is used for action ongoing in the (discourse-defined) present moment. No further functional exactitude is intended by the labels.

Table 6.5: Present continuous and present perfective formations in KRDS and pKmt

	Present definite formations: suffixed to verb stems			
	Present continuous			Present perfective
pKmt	*-ɛ-ʧʰ-AGR.I			*-i-ʧʰ-AGR.I
KS	-ɛ-ʧʰ-AGR.I			-i-ʧʰ-AGR.I
RL	-ɛ-s-AGR.I			-i-s-AGR.I
MH	-(ɛ)-ʧʰ-AGR.I			-i-ʧʰ-AGR.I
TH	-ʧʰ-AGR.I			-[i;e]-ʧʰ-AGR.I
RP			-bær næk-s-AGR.I	-s-AGR.I [a]
SH	-ɛ-s-AGR.IA			-i-ʃ-AGR.IB
BH			-ir ɖ̪ʱɔr-s-AGR.IB	-s-AGR.IB
BN		-ia as-AGR.IA	-iba lag-i-s-AGR.IB	-i-s-AGR.IB

[a] For an alternative description of this BH construction as 'inceptive', distinct from 'perfective', see Joshy and Joshy (2007:22, 25ff.).

The two present 'definite' formations shown in Table 6.5 are erstwhile compounds involving the auxiliary /aʧʰ-/ 'be present'. The auxiliary verb was grammaticalised as part of this construction and reduced to *-ʧʰ 'present tense', though it also remains in Magadhan lects as an independent and irregular verb. This grammaticalised piece of verbal morphology has been inherited into the Magadhan lects, though subsequently lost in Magahi and Bhojpuri (Chatterji 1926:1035). Its occurrence in proto Kamta is a retention.

The agreement system used in present 'definite' formations is AGR.I. This system of endings attaches: (a) directly to verb stems in the present indefinite formation (see §6.2.3); and (b) to the present tense marker, which is derived from the auxiliary verb *aʧʰ-. Diachronically, this distribution is explained if the AGR.I system was in effect prior to the grammaticalisation of *aʧʰ- > *-ʧʰ. At that time, the AGR.I system attached to verb stems, including *aʧʰ- 'be present'. After the auxiliary verb *aʧʰ- was grammaticalised as *-ʧʰ 'PRS' the same agreement endings were retained despite the change in morpho-syntactic environment. In recognition that the AGR.I system is an old and inherited system, it is termed the 'primary' system of agreement in IA studies.

The primary system is distinguished from the 'secondary' systems, which are: (a) later developments (not inherited from pre-proto Magadhan); and (b) attach not to the verb stem, but to erstwhile participial morphology reanalysed as tense morphology. Participial suffixes became a source of tense and aspect morphology during late MIA. The reanalysis of various participial morphemes as past and future tense markers is a proto Magadhan change, and discussed in §6.2.5 and §6.2.6. After the division of proto Magadhan, the secondary systems of agreement were innovated independently in proto Bangla, proto Asamiya and proto Kamta (see under §6.4 below). For this reason, the secondary systems of eastern Magadhan lects are considerably more differentiated than the primary systems.

The data in Table 6.5 show minor phonological changes in the inherited present tense marker, as well as phonological and morphological changes in the perfective and continuous aspectual markers.

Phonologically regular reflexes of *-ʧʰ ‘PRS’ occur in seven of the eight lects (cf. the intervocalic reflex of *ʧʰ in Table 4.8). The one exception is SH /ʃ/ in the present perfective conjugation; the anticipated reflex in SH is /s/ (see the present-continuous conjugation). This postalveleorisation (s > ʃ) in SH of the present tense morpheme is restricted to the perfective conjugations: the present perfective, and the past perfective (see §6.2.5). This change is morphologically and phonologically conditioned, and as a result of this complexity it is considered diagnostic of a propagation event. It is most plausible that the phoneme first underwent alveolarisation and fricativatisation *ʧʰ > s (the regular reflex), followed by morphologically-conditioned re-phonemicisation > /ʃ/.

[MI 53.] *-ʧʰ ‘PRS’ (>*-s) > /-ʃ/ ‘PRS’ in present and past perfective formations {SH}. Diagnostic.

Turning to the perfective-marking vowel in the present perfective formation, we find that:

- /-i/ is found across RL, MH, SH and BN;
- /-e/ is found in TH, with allomorph [-i] due to regressive raising; and
- no vowel (i.e. a zero marker) is found in RP and BH.

Given the broader range of perfective /i/, the TH form /-e/ can be viewed in two complementary ways. Firstly, the form /-e/ in TH simple verbs is an analogical extension of the perfective /-e/ in compound verbs which came from *-ia (see the arrow in Table 6.2). Secondly, SCB has undergone the same analogical extension, and the TH change represents a convergence towards the norms of SCB. The similarity between conjugations in TH and SCB extends also to the present continuous formations shown in Table 6.5. The TH present continuous is formed by suffixing the verb with the present tense marker—without aspectual marking—followed by the agreement endings. Other KRDS lects have some overt continuous aspectual marker in the corresponding formation, even if it is variable. These two changes in TH are diagnostic of contact relations between TH and SCB.

[MI 54.] *VERB-i-ʧʰ-AGR.I ‘present perfective’ replaced by VERB-e-ʧʰ-AGR.I ‘present perfective’ {TH}. Diagnostic of contact relations with SCB.

[MI 55.] *VERB-ɛ-ʧʰ-AGR.I ‘present continuous’ replaced by VERB-ʧʰ-AGR.I ‘present continuous’ {TH}. Diagnostic of contact relations with SCB.

Turning to the present perfective conjugation in RP and BH, the absence here of perfective *-i is plausibly associated with the phonological changes reconstructed in §4.4.6 (transposition and loss). However, the formulation of changes there requires that one of the adjacent consonants to the vowel be a sonorant. This condition is not always met in the present perfective conjugations, and yet the medial *i is still lost, for example in RP /dekʰ-s-o˜/ ‘I have seen’. The loss of medial *i is not predicted in this environment by the general phonological changes, and a morphologically-specific change must instead by proposed. What seems to have happened is that the medial *-i ‘PFV’ became zero in some verbal constructions by the phonological changes of transposition and loss of medial high vowels. The zero marker was then reanalysed as the regular marker of perfective in the present perfect conjugation.

[MI 56.] *VERB-i-s-AGR.I ‘present perfect’ > VERB^{H}-s-AGR.I {RP} and VERB^{H}-s-AGR.IA {BH} (after [MI 47.] and [PI 33.]). Diagnostic.[3]

---

[3] VERB^{H} indicates a verb with the last vowel of its stem raised *ɔ > [o], *ɛ > [e].The raising does not apply to *a, unlike in SCB.

There is a small difference between BH and RP in agreement endings in this construction, to be discussed in §6.4.1.1. Despite this difference, the reanalysis of zero as the perfective marking is common to both RP and BH. It is unique, and the range is sociohistorically plausible as a zone of integrated propagation, and therefore the change is diagnostic of a propagation event. The difference in agreement endings shows that the change [MI 56.] must have occurred subsequent to Progressive Vowel Raising ([PI 23.]). Furthermore, as this change is plausibly a reanalysis of the zero allomorph created by transposition and loss of medial vowels, therefore [MI 56.] must also have occurred after [PI 33.].

Based on this reconstruction of changes, the past perfective formation inherited from proto Kamta was: *VERB-i-ʧʰ-AGR.I

The last of the present tense formations to be discussed is the present continuous, with the reconstruction given in Table 6.5 as *VERB-ɛ-ʧʰ-AGR.I. The present continuous constructions in KS, RL, MH and SH are all very similar (see Table 6.5) and may be straightforwardly reconstructed to a common formation. In contrast, the present continuous formations in RP, BH and BN are noncognate and periphrastic. These are very likely to be recent innovations through the reanalysis of certain compound verbs with continuous aspect (see further below). The TH form for present continuous lacks any overt aspectual marking and has been reconstructed as an innovation in [MI 55.]. This leaves only the formation *VERB-ɛ-ʧʰ-AGR.I as a possible candidate for retention from the proto Kamta stage. If this formation was found only in KS, RL and MH we might consider it a recent innovation, as these three lects form a historical subgroup and have undergone common changes after the division of proto Kamta (cf. §7.5.2). However, the same construction is also found in SH. There is no diagnostic innovation whose range involves western KRDS lects and SH while excluding the other central KRDS lects—RP and BH. The distribution of this tense formation in both western KRDS lects {KS, RL, MH} as well as in the central KRDS lect {SH} justifies its reconstruction as a proto Kamta inheritance.

A plausible cognate for the proto Kamta aspectual marker *-ɛ is the Maithili continuous morpheme /-ʌi/. If the sequence *ɔi (where *ɔ corresponds to Maithili /ʌ/) had been inherited into KRDS as a morpheme with continuous function, then the phonological reflex could very plausibly have been *-ɛ. The diachronic phonological plausibility comes from considering that the proto Kamta third person agreement endings *-ɛ are the reflex of Magadhan *-ai* > *ɔi. Another etymological possibility concerns the verbal nominaliser /-a/ (possibly passive participle, see Chatterji 1926:660) suffixed with the old locative *-ɛ, followed by vowel coalescence: *-aɛ > *-ɛ. Further reconstruction at a broader Magadhan level is required before the etymology can be finalised.

Turning to the periphrastic present continuous constructions in Table 6.5, they are:

- RP: -bær næk-s-AGR.I
- BH: -ir ɖ̤ʱɔr-s-AGR.IB
- BN: -iba lag-i-s-AGR.IB

The two elements that make up these formations are (1) the infinitive (cf. §6.2.2) and (2) an auxiliary verb (either reflexes of proto Kamta *lag- ‘attach’ or *ɖ̤ʱɔr- ‘catch’) grammaticalised with continuous aspect. Three diachronic interpretations are possible regarding the differences in these formations. Firstly, a common change may have been propagated across all three lects whereby a compound construction *[VERB-INF lag-PFV-PRS-AGR] was reanalysed with continuous aspect. This would then have been followed by

the substitution of the verb *lag ‘attach’ with *ɖ̪ʱɔr- ‘catch’ in BH only, but still with continuous aspect. This is a plausible enough interpretation, and gives weight to the close phylogenetic relations between RP and BH lects—or to use more general labels, the lects of Rangpur and Cooch Behar regions. The second possible interpretation of the three periphrastic continuous formations is that RP and BN constitute a common change (based on the cognacy of the grammaticalised auxiliary verb < *lag- ‘attach’). The BH formation would then be considered an independent change. This interpretation gives weight to the cognacy of the contemporary forms in BN and RP. It is less sociohistorically plausible than the first interpretation because it does not recognise the consistently closer phylogenetic relationship between RP and BH than between RP and BN. Furthermore, the similarity of a periphrastic continuous formation in both RP and BH is left unexplained by this second interpretation. The third interpretation is that all three lects independently innovated periphrastic continuous aspect formations. These three interpretations all account for the divergences in the data. I hold that the first explanation is the more plausible (because of the closer phylogenetic relations between BH and RP), and therefore reconstruct the following common change for RP, BH and BN:

[MI 57.] VERB-INF + present perfective of *lag- ‘attach’ > ‘present continuous’ {RP, BH, BN}. Supportive, not diagnostic.

This change involves the propagation of a reanalysed meaning for a construction already present in the lects. The precise forms that made up the construction, while cognate, need not have been identical at the time of propagation. That is, it is not necessary that [MI 57.] occurred before (a) the change of initial *l > n in RP and BH or (b) the reanalysis of the infinitive in those lects by [MI 51.]. The change [MI 57.] is somewhat complex in its conditioning, but because the change is generalised for RP, BH and BN based on sociohistorical plausibility (see above), the change cannot in turn be used to diagnose sociohistorical relations.

Finally, BN has a second present continuous formation: VERB-ia as-AGR.IA. As a periphrastic construction, this is also likely to be a recent innovation. The morphemes involved are the perfective /-ia/ and /as-/ ‘be present’ from which we get the grammaticalised present tense marker *-ʧʰ. (Recall that the verb *aʧʰ- persists across KRDS as an independent lexeme alongside the grammaticalised morpheme *-ʧʰ.)

6.2.5 Past tense constructions

The constructions examined in this section are labelled ‘past tense’; this differs from Masica’s classification of NIA tense/aspect categories. Masica holds that the -l- form is an *unspecified perfective* across New Indo-Aryan, and including SCB and SCA—the near relatives of KRDS. Regardless of whether this perfective label is appropriate for SCA and SCB, it does not seem appropriate for the KRDS situation. In KRDS if any formation is to be considered a candidate for an *unspecified* perfective, it should be the VERB-i-ʧʰ-AGR.I construction labelled above as ‘present perfective’. Unlike the formation VERB-il-AGR.IIA which only occurs for past tense in KRDS, the /-i-ʧʰ-/ formation, while generally present perfective, can be used as past perfective given the right discourse context. The label ‘past tense’ for the /-l/ suffix is thus more suitable than Masica’s terminology in the case of KRDS.

Throughout Table 6.6 [l,n] alternation is conditioned by the phonology of the agreement protosuffixes: the /-n-/ allomorph occuring before first person endings (extended to some second person endings in RP); the /-l-/ allomorph elsewhere (see further §6.4.1.3). For example (from MH):

- /mui baʤar gɛi-**n-u**/ < *mui baʤar gɛi-**l-ũ̯**
  'I bazaar go-PST-1SG' = 'I went to the bazaar'
- /t̪ui baʤar gɛi-**l-o**/ < *t̪ui baʤar gɛi-**l-o**
  'you bazaar go-PST-2SG' = 'You went to the bazaar'

Table 6.6: Past tense formations in KRDS and pKmt

	Past indefinite (unspecified aspect)	Past definite		
		Past continuous		Past perfective
pKmt	*-il-AGR.IIA			*-i ʧʰil-AGR.IIC
RL	-[l;n]-AGR.IIA			-isi[l;n]-AGR.IIA
MH	-[l;n]-AGR.IIA			-is[l;n]-AGR.IIA
TH	-i[l;n]-AGR.IIA		-ʧʰi[l;n]-AGR.IIC	-iʃ[l;n]-AGR.IIC
RP	-(i)[l;n]-AGR.IIA	-bær næk-si[l;n]-AGR.IIC		-si[l;n]-AGR.IIC
SH	-i[l;n]-AGR.IIA			-iʃ[l;n]-AGR.IIC
BH	-(i)l-AGR.IIA	-ir ɖ̪ʱɔr-sil-AGR.IIC	-ia tsʰil-AGR.IIA	-i sil-AGR.IIC
BN	-il-AGR.IIA		-ia asil-AGR.IIA	-i sil-AGR.IIA

As in the present tense, the past tense formations are quite uniform in the perfective aspect. The past indefinite formations (with unspecified aspect) are also very uniform across the eight lects. It is once again the continuous aspectual formations which lack uniformity across the lects.

The proto Kamta simple past is reconstructed as *-il. The /-n/ element is a more recent innovation—diachronically conditioned by the nasalised vowels in the first person AGR endings (see §6.4.1). The reconstructed morpheme *-il 'past tense' is cognate with Oriya, Bangla and Asamiya (i.e. eastern Magadhan lects), and partially cognate also with the western and central Magadhan lects which have /-ʌl;-əl/. This morpheme is not inherited from OIA participial or tense morphology, and its etymology beyond MIA is somewhat tricky (cf. Chatterji 1926:940ff.). Chatterji notes that the MIA phonological changes had eroded the inherited OIA passive participle *-ta; ita* > *-a; ia* to the point of being non-distinctive. Based on the presence of the /-l/ element in Magadhan languages he surmises that, during the common Magadhan stage (the Magadhi Apabhraṁśa), the inherited and eroded passive participle was extended by *-ll* > *-ila, ala*. Changes in verbal syntax between MIA and NIA—attested in written records—account for the altered function of the passive participle, as described in the following quote (ibid.:939–940).

> [T]he passive participle construction, the verb being an adjective qualifying the nominative when it was intransitive and the object when transitive, became the common idiomatic way of expressing the past in MIA. By the time when the Apabhraṁśa Stage came in, the old inflected past forms, which still lingered in Second MIA., were clean swept away, and only this participial past remained in IA.; and the NIA. past tense was formed out of this.

This hypothesis accounts for the presence of *-l-* based past (or perfective) morphemes in the Magadhan languages, and the transition from passive construction to active, with the accompanying creation of secondary systems of agreement.

Turning our attention back to the past formations of Table 6.6, the various sibilants and affricates in past definite formations are traced to a grammaticalised form of the verb *atʃʰ- 'be present' (as was the case in the present tense cf. §6.2.4). In these data, TH has deaffricated the morpheme *-tʃʰ > -ʃ, but the conditioning is different to that found in the SH present tense formations. In this TH formation, the protophoneme *-tʃʰ has moved next to another consonant by the loss of the intervening vowel. The phonological sequence that results is phonotactically disfavoured, with no examples of /tʃʰC/ in the TH data. Accordingly, the affricate has been deaffricated to give the more phonotactically favourable sequence /ʃC/. The divergence in the past continuous formation in TH is therefore a phonological, rather than a morphological change. It is of little complexity and has no value for diagnosing propagation events.

The most variety in the past tense formations, as in the present tense, comes in continuous constructions. With no specifically past continuous formation found in the lects RL, MH and SH, and innovative periphrastic constructions in RP and BH, there is insufficient evidence to reconstruct a proto kamta past continuous distinct from the past perfective. This slot is accordingly left blank in the bottom row of Table 6.6 as well as in Table 6.1 earlier in the chapter. Forms to distinguish past continuous from past perfective function are post-proto Kamta innovations, and accordingly have quite localised ranges.

The past continuous formation in RP and BH is the past tense analogue of the construction innovated in the present tense formations by [MI 57.]. The periphrastic continuous based on the auxiliary *lag- is not part of the data collected at BN, though this probably reflects an inadequacy in the data rather than the absence of the construction in the lect. The TH construction VERB-tʃʰi[l;n]-AGR.IIC is once again identical with Bangla, and is reconstructed as a morphological replacement, diagnostic of contact relations.

[MI 58.] > VERB-tʃʰi[l;n]-AGR.IIC 'present continuous' {TH}. Diagnostic of contact relations with SCB.

The other past continuous formation—found in both BH and BN—is a compound verb construction, but in this case the vector verb is *atʃʰ- 'be present', and the main verb is suffixed with the perfective participle *-ia. A similar construction was seen in the present-continuous conjugation in BN. The structure of this construction is the same as used in Asamiya for a disambiguated past continuous function. There is a slight difference in form because the perfective participle in the BH and BN constructions is /-ia/ and in the Asamiya construction it is /-i/. Nonetheless, contact with Asamiya is a likely conditioning factor for the range of propagation of this construction—BN being within Assam, and BH located very near the border with Assam. On the other hand, there is some evidence elsewhere to support a propagation event connecting BH and BN (see the initial devoicing change in Chapter 4). Lacking clear reason to decide between these two possible explanations—contact with Asamiya, versus propagation between BN and BH—the change [MI 59.] is stated as having an ambiguous diagnostic value.

[MR 59.] > VERB-PFV *atʃʰ-il-AGR.IIA 'past continuous' {BN, BH}. Diagnostic value ambiguous between contact relations with Asamiya or a PE within BH and BN.

This concludes the discussion and reconstruction of past tense formations for proto Kamta.

6.2.6 Future tense constructions

Only one future tense construction was collected as part of this study, though, as for the other tenses, it is possible that further categories of *Aktionsart* are possible by compounding with auxiliary verbs. Analogously to the [l,n] altenation in the past tense morphology, [b,m] alternation in Table 6.7 is conditioned by the historical phonology of the agreement suffixes. The /-im/ allomorph occurs before first person agreement endings and is subsequently extended in RP to the 2PL ending; the /-ib/ allomorph is found before other endings (see further §6.4.1.3).

Table 6.7: Future tense formations in KRDS and pKmt

	Simple future
pKmt	*-i[b; m] -AGR.IIB
KS	-[b; im] -AGR.IIB
RL	-[b; im] -AGR.IIB
MH	-[b; im] -AGR.IIB
TH	-[ib; im] -AGR.IIB
RP	-[(i)b; im] -AGR.IIB
SH	-[ib; im] -AGR.IIB
BH	-[ib; im] -AGR.IIB
BN	-[ib; im] -AGR.IIB

The history of this formation is straightforwardly reconstructed. The element /i/ is lost in some lects, but maintained throughout KRDS in the fused future-and-first-person-singular-ending /-im/; it thus constitutes part of the inherited material. The [-m] allomorph is the historical result of transfering the nasalisation from a suffixed vowel onto the tense marker (cf. [MI 65.]). This morphologically conditioned nasalisation of *-ib 'FUT' occurred prior to proto Kamta and thus the reconstructed proro Kamta system includes both allomorphs *-[ib; im].

The future tense employs a partially distinct set of agreement suffixes (AGR.IIB) to the past formations; these are described and reconstructed in §6.3–§6.4.

Similarly to the past tense marker, the future tense marker *-ib is a reflex of older participial morphology: 'it comes from the OIA. future passive participle gerundive in «-tavya-» or «itavya», in second MIA. «-(i)avva-, -(i)abba-, -ĕbba »' (Chatterji 1926:965). As in the case of the past tense morpheme *-il, the use of the morpheme *-ib for *active* future tense is a late MIA or early NIA innovation. Only after this morpheme started to be used in active constructions were secondary endings added to the future tense formation. The AGR.IIA (past) and AGR.IIB (future) systems thus constitute early NIA innovations, and are relevant to subgrouping within eastern Magadhan (see further §6.4).

6.3 Agreement endings: synchronic description

This section sets out the verbal systems of agreement for each of the eight KRDS test sites. The agreement is almost entirely between the verb and the grammatical subject. The only exception is the system of patient agreement found in central Jhapa and Morangia Rajbanshi (see Rangeli, §6.3.2) which has been described in detail by Wilde (2008:§3.7).

Peculiar features of the agreement systems in each lect are noted in passing. The proto Kamta system of agreement marking is reconstructed in §6.4.

Agreement marking on the verb is a common feature of NIA. In most NIA lects, the agreement is with the subject of the clause. In some languages there is additional marking for the object, for example Maithili, and as mentioned above, Rajbanshi in some parts of Nepal). The notion of subject is defined differently across grammatical traditions. Here the intended referent is the S or A of a finite clause (Comrie 1978). This constituent receives no overt nominal marking in KRDS (see §5.3.3), but the person and number of the subject control a morphological position on the verb. Agreement is only characteristic of KRDS finite verbs, and is absent in nonfinite verbs.

In KRDS lects there are multiple sets of agreement endings, each used with particular tenses. (For the broader NIA perspective, see Masica 1991:259ff.) Agreement endings are traditionally classed in IA studies as Primary and Secondary. This classification is basically diachronic, and not the same as Wilde's (synchronic) distinction between primary agreement (with the agent) and secondary agreement (with the patient). As this study is historically oriented, and patient agreement is restricted to just one of the eight sample lects, the terms 'primary' and 'secondary' agreement are henceforth used with their traditional IA meanings.

As outlined in §6.2.4, the primary system of agreement is inherited from MIA, and thus historically antecedent to the secondary system, which is a NIA innovation (see further §6.4). There is also a synchronic pattern to these diachronically distinct systems: the primary endings (AGR.I) attach to present tense formations; the secondary (AGR.II) endings attach to past and future tense formations. The position of the imperative endings (AGR.IMP) is alike to the primary endings—directly after the verb stem—while the function of the clause differs according to which set of endings is used (but see §6.2.3 on the need for further synchronic analysis).

Amongst the secondary endings, there are further divisions of labour, with one set of endings (AGR.IIA) used with the past tense /-(i)l-/ and a partially distinct set of endings with the future tense /-(i)b-/ (AGR.IIB). Additionally, in some KRDS lects the third person endings differ depending on whether the past tense is a 'definite' past or a simple past formation, thus yielding a third subset of secondary endings: AGR.IIC (the historical explanation for AGR.IIC is in §6.4.2).

I have stated above that agreement in KRDS involves both the person and number of the subject. This finding is notable because several IA studies state that number marking is absent from the Bangla-Asamiya subgroup, for example:

> Note the neutralization of number in the 1st person … in the Eastern languages (except Oriya) as far west as Bhojpuri, … The distinctions remaining in the 2nd and 3rd persons … in the east have become distinctions of honorificity rather than of number. (Masica 1991:285, n.7.)

The secondary agreement system of KRDS—in particular the use of cognate forms to mark singular and plural categories of number—is a central proof for the proto Kamta stage of development. Number is not marked in the secondary systems of Bangla or Asamiya, and hence its presence in KRDS takes on added significance: it provides proof for a stage of KRDS linguistic history which is distinct from that of Bangla and Asamiya (cf. §6.4.1). For a detailed reconstruction of the history of number marking in KRDS and eastern Magadhan see §6.4.

6.3.1 Kishanganj (KS)

The agreement system used in and around Kishanganj (KS) of Bihar is outlined in Table 6.8.

Table 6.8: Subject agreement system for Kishanganj

	1	2	3
AGR.IMP	-i	-ʌ	[a]
AGR.I	-i	-is	-ɛ
AGR.IIA	-u	-o	-ɛ
AGR.IIB	-Ø	*as AGR.IIA*	

[a] Data missing.

The KS agreement system differs from the rest of KRDS (though not BN) by the absence of number marking.

6.3.2 Rangeli (RL)

The complexity of agreement in Rangeli (RL) of Nepal sets this lect apart from the other KRDS lects included in this study. RL has one system of endings for agreement with the subject only, and one system for agreement with the patient or other non-nominative nominal. These two systems of endings are summarised in Tables 6.9–6.10. The data in Table 6.10 are drawn from Wilde (2008:412–414).

Table 6.9: RL system of agreement, with subject only

	1SG	1PL	2SG	2PL	3
AGR.IMP	-u	-i	-[ɛ]k	-ʌ	-ok
AGR.I	-u	-i	-i	-ʌn	-ɛ
AGR.IIA	-u	-ʌ	-o	-ʌn	-ɛ [a]
AGR.IIB	-Ø	*as AGR.IIA*			

[a] Wilde (ibid.:411) has this morpheme as [ɪ].

Unlike KS which lacks number marking, in RL the subject-only agreement system distinguishes singular and plural for both first and second persons. The categories of inflection are slightly more complex than just singular versus plural. As for the pronouns (see §5.5.2), the plural may also be used as an honorific singular. This extension of plural meaning to cover honorific singular is common in NIA. The two second person categories are labelled as 2SG and 2PL because these are the principle categories of the system, not honour.

Together with MH, the RL system is unique in KRDS for employing a second person plural ending based on the vowel *ɔ > ʌ {MH, RL}, rather than *ɛ. This variation is examined in §6.4.2.

Of the eight KRDS lects examined here, RL is the only one with patient agreement as well as subject agreement. As this feature is only found in one of the eight sample lects it is not useful for broad subgrouping within KRDS. It does, however, show the distinctiveness of RL, and may furthermore be diagnostic of historical relations between RL and Maithili.

Detailed analysis of patient agreement may be found in Wilde, with some discussion of its (possible) origins through contact with Maithili (2008:§3.7).

Table 6.10: RL system of agreement, with patient marking

		Patient			
		1	2SG	2PL	3
AGR.Subjunctive and AGR.Past habitual [a]					
Agent	1	–	-ku		-ki
	2	-is	–		-is
	3	-ɛ	-ku		-ki
AGR.I					
Agent	1	–	-uŋgu		-iŋgi
	2	-i	–		-i
	3	-ɛ	-u		-i
AGR.IIA					
Agent	1	–	-gu		-gi
	2	-o	–		-o, -gi
	3	-ɪ	-gu		-gi
AGR.IIB					
Agent	1	–	-gu		-gi
	2	-o	–		-o
	3	-ɪ	-t̪-u-ku [b]	-t̪-ʌ-kʌn [b]	-gi

[a] As the data in this table are drawn from Wilde (2008), I have maintained his verbal category of 'subjunctive' so as to maintain the integrity of the data.

[b] The morpheme /-t̪/ may be a fusion of future tense marking and agreement.

6.3.3 Mahayespur (MH)

The MH agreement endings outlined in Table 6.11 are highly similar to that of RL in Table 6.9.

Table 6.11: Subject agreement system for Mahayespur

	1SG	1PL	2SG	2PL	3
AGR.IMP	-u	-i	-ɛk	-ʌ, -[ʌ]n	-ok
AGR.I	-u	-i	-is, -i	-ʌn	-ɛ
AGR.IIA	-u	-ʌ	-o	-ʌn	-Ø, -ɛ
AGR.IIB	-Ø	*as AGR.IIA*			-ɛ

There are two forms categorised as '2SG' in both the primary and imperative systems. Speakers do not give a consistent explanation of the semantic or functional difference between these variants, and closer synchronic study of texts is requried. The variation in the primary form [-is, -i] '2SG' is a phonological variation in word-final /s/. Variation in the third person past (AGR.IIA) endings between zero and /-ɛ/ is yet to be shown to reflect any grammatical distinction.

6.3.4 Thakurgaon (TH)

The agreement data for the TH test site are in Table 6.12:

Table 6.12: Subject agreement system for Thakurgaon

	1SG	1PL	2SG	2PL	3
AGR.IMP	-u	-i	-ek	-[e]n	-ok
AGR.I	-u	-i	-i	-en	-e
AGR.IIA	-u	-o	-o	-en	-Ø, -e
AGR.IIB	-Ø	*as AGR.IIA*			-e
AGR.IIC	*as AGR.IIA*				-o

The TH system has one structural difference to that of MH above: the third person AGR.IIA (past tense) has distinct endings in past indefinite and past definite conjugations. For example:

- /d̪exil**e**, d̪exil/ ‘she saw’. Past indefinite.
- /d̪exiʃl**o**/ ‘she had seen’. Past perfect.
- /d̪ɛxʧʰil**o**/ ‘she was seeing’. Past continuous.

The past indefinite conjugation uses the AGR.IIA ending, while the past definite conjugations take the AGR.IIC ending. Verbal transitivity may be relevant to third person agreement marking in TH, but it remains to be studied. The historical origin of the AGR.IIC system is outlined in §6.4.2.3.

The TH system is akin to the other systems below, and distinguished from MH and RL above by having second person plural endings based on *ɛ (> /e/) instead of *ɔ. Interestingly, both *-ɛ (> -e) and *-ɔ (> -o) are found as third person markers in TH—a crucial fact in the reconstruction of second person plural markers *-[ɛn;ɔn] (see §6.4.2).

6.3.5 Shalkumar (SH)

The agreement endings for SH are given in Table 6.13:

Table 6.13: Subject agreement system for Shalkumar

	1SG	1PL	2SG	2PL	3
AGR.IMP	-o	-i	[4]	-[e]n	-[u]k
AGR.IA	-o	-i	-i̪t	-en	-e
AGR.IB	-u	*as AGR.IA*			
AGR.IIA	-u	-i	-o	-en	-ek
AGR.IIB	-Ø	-o	-o	-en	-e
AGR.IIC	*as AGR.IIA*		-i̪t	-en	-o

[4] Data missing.

The SH agreement system is structurally somewhat different to the systems above. As in TH there is a difference in conjugation between past definite and past indefinite formations in the third person endings. In SH this split in conjugation in the past formations is also extended to the second person endings, thus:

/ḍ̪exil-o/ 2SG past **indefinite** for the verb /ḍ̪ɛk^{h}-/ ‘see’
/ḍ̪exiʃil-iṭ̪/ 2SG past **definite** for the verb /ḍ̪ɛk^{h}-/ ‘see’

The impact of verbal transitivity on agreement marking in SH remains to be studied.

The SH system also differs from those described above in the primary endings. The first person singular ending in the present *perfective* is different to that found in the present *indefinite* and *continuous* conjugations:

/ḍ̪ex-i-s-**u**/ 1SG present **perfective** for the verb /ḍ̪ɛk^{h}-/ ‘see’
/ḍ̪ex-ɛ-s-**o**/ 1SG present **continuous** for the verb /ḍ̪ɛk^{h}-/ ‘see’
/ḍ̪ex-**o**/ 1SG present **indefinite** for the verb /ḍ̪ɛk^{h}-/ ‘see’

Progressive raising of *o > /u/ is not a phonologically regular feature of SH, and hence this seems to be a morphologically-specific raising process.

This description points to a further difference between SH and the preceding systems: in both AGR.IA and AGR.IMP the first person singular endings are a lower vowel /-o/, in contrast with the ending /-u/ found in the western KRDS systems described above. An historical explanation for this difference is given in §6.4.1.

Finally, the SH system extends the pleonastic /-k/ found in imperatives to the third person past ending (as does BH below).

6.3.6 Rangpur (RP)

The Rangpur system shares the same overall structure as TH.

Table 6.14: Subject agreement system for Rangpur

	1SG	1PL	2SG	2PL	3
AGR.IMP	-o˜	-i	-ek	-o, -[e]n	-uk
AGR.I	-o˜	-i	-iʃ	-en	-e
AGR.IIA	-u	-o	-u	-en	-Ø
AGR.IIB	-Ø	*as AGR.IIA*			-e
AGR.IIC	*as AGR.IIA*				-o

The RP system has some phonological differences with the foregoing systems: firstly, RP maintains the inherited nasalisation on first person singular endings; secondly, the vowel in the second person singular ending of the secondary systems /-u/ is higher than for the lects described above. The raised vowel is the result of progressive vowel harmony (see §4.4.2).

6.3.7 Bhatibari (BH)

The Bhatibari system shares the structure of RP and TH in distinguishing the AGR.IIC system. The endings are displayed in Table 6.15.

Table 6.15: Subject agreement system for Bhatibari

	1SG	1PL	2SG	2PL	3
AGR.IMP	-[o]ŋ	-i	-Ø, -ek	-o	-uk
AGR.IA	-[o]ŋ	-i	-iʃ	-en	-e
AGR.IB	-[u]ŋ	*as AGR.IA*			
AGR.IIA	-uŋ	-oŋ	-u	-en	-ek
AGR.IIB	-Ø	-o	*as AGR.IIA*		-e
AGR.IIC	*as AGR.IIA*				-o

There is one structural difference between this system and those above: the split of primary endings into two systems AGR.IA and AGR.IB. These two primary systems differ only with respect to first person singular: in the AGR.IB system the ending has been raised to /-uŋ/ from /-oŋ/. Some historical explanation for the raising of AGR.IB /-uŋ/ has already been given in §6.2.4.

Similarly to RP, BH has progressive vowel harmony, which affects the height of vowels in Table 6.15, for example /-u/ '2SG'.

6.3.8 Bongaigaon (BN)

The system of subject agreement in BN is quite different from the general pattern outlined above for the other KRDS lects. Similarly to KS, agreement in BN is not inflected for the number of the subject.

Table 6.16: Subject agreement system for Bongaigaon

	1	2		3
		Low	High	
AGR.IMP	-o˜ $^{C\_}$, ŋ $^{V\_}$	-ɛn		-ɔk
AGR.I	-o˜ $^{C\_}$, ŋ $^{V\_}$	-is	-a	-ɛ
AGR.IIA	-o˜	-i	-a	-Ø
AGR.IIB	-Ø	-i		-ɔ

While number is unmarked in BN, high versus low honour is distinguished in the second person endings. The ending /-a/ is not found elsewhere in KRDS, and is shared instead with SCA.[5] The first person ending is the lower variant /-o˜/ found also in RP, SH and BH. The high variant /-u, -u˜/ is not found in BN, nor in SCA. The relationship of the BN and SCA agreement systems is close, and the history is reconstructed in §6.4.3.

6.4 Agreement endings: historical reconstruction

The specification of agreement endings with particular conjugations is residual from multiple layers of historical change which swept over the verbal morphology during the MIA and NIA periods. Of the various sets of agreement endings, the primary set alone constitutes

[5] But compare §7.3.1.3.

an historical continuation in NIA of the OIA agreement morphology (Masica 1991:260). As inherited features, cognacy of primary endings in NIA lects is generally not indicative of change events, and hence not useful for subgrouping. NIA primary endings are (mostly regular) reflexes of the Sanskrit present active endings. (Cf. Bubenik 2003:227–228. Regarding the regularity of the reflexes, see the note under Table 6.17 below)

Table 6.17: Derivation of primary endings in NIA from OIA and MIA forms

	1SG	2SG	3SG	1PL	2PL	3PL
Pāli (early MIA)	*-āmi*	*-asi*	*-ati*	*-āma*	*-atha*	*-anti*
Prakrit (middle MIA)	*-āmi*	*-asi*	*-ai*	*-āmo*	*-aha*	*-anti*
Apabhraṁśa (late MIA)	*-auṁ*	*-ahi*, *-asi*	*-ai*	*-ahuṁ*	*-ahu* [a]	*-ahiṁ*
various NIA languages (from Masica 1991:263–264)	-o˜, -u˜, -əu˜, -əo˜, etc.	-əi, -əe, -es, -iʃ, -əs, etc.	-y (i.e. -j), -e, -əe, -əi, etc.	-əu˜, -au˜, -u˜, -o˜ etc.	-ə(n), -ɔ, -o, -əo	-ɔnṭi, -ən, -in, -en, -e˜, etc.

[a] 'There are various difficulties connected with explaining the origin of the plural suffixes in Apabhraṁśa. *u* in *-ahu* in the second person, as suggested by Bloch, comes probably from the suffixes of the third person imperative *-a(h)u* (< *-atu*) and *-antu*' Bubenik (2003:228).

The outcome of this historical continuity is that *cognacy in primary endings is not unusual or unexpected in NIA languages*, and should not on the face of it be considered diagnostic of a (NIA) propagation event.

The situation is markedly different when we come to the secondary endings. The inherited perfective and future conjugations of OIA were lost during MIA and early NIA. These were replaced in the Magadhan lects by constructions based on passive participles, reanalysed firstly as past and future tense markers with passive voice and subsequently with an active sense (cf. §6.2.5 and §6.2.6). It was only after this morphosyntactic reanalysis that the secondary sets of agreement endings were created. As Chatterji puts it: 'Affixation for the participial tenses is a NIA. development' (1926:967). Given that these secondary sets of affixes were created *de novo*, cognacy among the Magadhan lects in secondary endings is not to be regarded as retention from OIA (or MIA for that matter). Instead, cognacy is diagnostic of *either proto Magadhan or post-proto Magadhan morphological innovations*.

In order to reconstruct the history of KRDS agreement marking and its origins in proto Magadhan, the eight KRDS systems are compared with each other as well as with corresponding forms for other Magadhan lects. Changes in the secondary systems are particularly significant for diagnosing propagation events given that these systems of endings are completely innovative.

With the reconstruction involving multiple morphemes in multiple systems, the discussion below becomes quite intricate. To assist the reader in following the details of the reconstruction, the final product of reconstruction—the agreement system of proto

Kamta—is reproduced at several points throughout the chapter with shading to indicate which morpheme is currently under discussion.

In the KRDS data, endings with first person function are etymologically distinct from endings for the other persons. The same cannot be said for the second and third person endings. In their case, the etymologies are so intertwined that the reconstruction procedes perspicuously only if second and third person endings are compared and reconstructed together. This perspicuity of analysis suggests that the endings for second and third person are reflexes of a pre-system which did not categorically distinguish these persons in agreement marking. The categories of this pre-system would have been first person versus second-third person general (2/3GEN). The latter category of inflection was then split, non-systematically, into second versus third person inflection. This hypothesis is discussed further in §6.4.2.

6.4.1 First person agreement suffixes

Table 6.18 presents the first person agreement suffixes for the eight KRDS sites. The data are divided into columns which indicate their functions within primary or secondary systems. Blank cells indicate that the category is absent from the relevant lect. Shaded cells are noncognate.

Table 6.18: First person agreement endings in KRDS and pKmt

	AGR.I (or AGR.IA)		AGR.IB		AGR.IIA		AGR.IIB		AGR.IIC	
	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL
pKmt	*-ɔw˜	*-i			*-u˜	*-ɔ˜	*-Ø	*-ɔ˜	< AGR.IIA	
KS		-i			-u		-Ø			
RL	-u	-i			-u	-ʌ	-Ø	-ʌ		
MH	-u	-i			-u	-ʌ	-Ø	-ʌ		
TH	-u	-i			-u	-o	-Ø	-o	-u	-o
SH	-o	-i	-u	-i	-u	-i	-Ø	-o	-u	-i
			< AGR.IIA							
RP	-o˜	-i			-u	-o	-Ø	-o	-u	-o
BH	-[o]ŋ	-i	-[u]ŋ	-i	-uŋ	-oŋ	-Ø	-o	-uŋ	-oŋ
			< AGR.I							
BN	-o˜ $^{C\_}$, ŋ $^{V\_}$				-o˜		-Ø			

The reconstructed first person endings, and the changes they entail, are discussed and argued for below, moving from left to right through the columns of Table 6.18.

6.4.1.1 Primary endings

KRDS primary endings for first person *singular* are divided between RL, MH and TH (which are /-u/) and SH, RP, BH and BN (which are /-o/). This divergence has been explained by phonological reconstruction of the protosequence *ɔw in §4.4.9. This protosequence goes to /u/ in the west and /o/ in the centre and east of KRDS. The only noncognate first person singular ending in Table 6.18 is KS /-i/, which comes from the first person plural (in the primary system). A consistent feature of the KS system is the absence

of number as a category of agreement. It will be argued below that KS inherited the number distinction from proto Kamta, but lost the distinction by reinterpreting the functions of inherited morphemes (see §6.4.3).

The RL agreement morphology differs from the other KRDS lects in many ways in both the primary and secondary systems due to the introduction of patient agreement. This patient agreement is not inherited and constitutes an innovation of RL (with central Jhapa, see Wilde 2008:§3.7). The multiple changes involved have not been reconstructed in detail, but are grouped together as changes which introduce a system of patient agreement. These changes are linguistically complex and are diagnostic of RL as a propagation network. The similarities and differences between the RL and central Jhapa systems of patient agreement need to be studied before they can be subgrouped based on this innovation.

[MI 60.] Introduction of patient agreement in verbal morphology {RL}. Diagnostic.

The primary first person *plural* endings are remarkably stable, and the inherited form is reconstructed as *-i. The only exception to this is BN, which, similarly to KS (though geographically on opposite sides of the KRDS area) lacks number marking in its agreement morphology. As in the case of KS, the historical implications of the absence of number marking in BN are considered at the end of this reconstruction (§6.4.3).

The first person plural ending *-i is not a reflex of the functionally corresponding MIA form *-āma* shown in Table 6.17. Plausibly cognate endings are also found in the primary agreement systems of SCB, Bhojpuri and Maithili. Tiwari (1960:167) notes that Old Bhojpuri distinguished number in its agreement morphology (Modern Bhojpuri does not), and that the primary first person endings were: -oˇ ‘1SG’, -īˇ ‘1PL’. He reconstructs the 1SG ending in Old Bhojpuri -oˇ as a reflex of Sanskrit *-ami* (> MIA *-auṁ*, from which KRDS -ɔwˇ is derived). Tiwari further proposes that the old Bhojpuri first person plural ending is a reflex of OIA suffix *-yaṭe* > -iaï > ie > ī (with the nasalisation a Bhojpuri innovation). Chatterji (1926:864) lists this OIA suffix *-yaṭe* with passive function in the OIA indicative present conjugation.

Whether this latter etymology of Tiwari’s is permitted to stand or not, the distribution of *-i in the first person primary endings of old Bhojpuri, as well as early Maithili (Jha 1985 [1958]:480) shows that this morpheme was inherited as part of the primary system since at least proto Magadhan—with reflexes in both western Magadhan (Bhojpuri), central Magadhan (Maithili) and eastern Magadhan (Bangla, KRDS). Therefore inheritance of both *-ɔwˇ and *-i from proto Magadhan into KRDS does not entail any post-Magadhan subgrouping relations for KRDS.

Reproduction of **Table 6.25:** Reconstruced pKmt system of AGR endings

	AGR.I	AGR.IMP	AGR.IIA	AGR.IIB	AGR.IIC
1SG	*-ɔwˇ		*-uˇ	*-Ø < *-uˇ	*-uˇ
1PL	*-i			*-ɔˇ	
2SG	*-iʃ	*-ɛkɔ		*-o	
2PL	*-[ɔ;ɛ]n(ṭi)	*-ɔ		*-[ɔ;ɛ]n(ṭi)	
3	*-ɛ	*-[o;u]kɔ	*[-Ø; -ɛ]	*-ɛ	*-ɔ

Returning to the KRDS data, the first person singular endings in the AGR.IB (present perfective) system of SH and BH are /-u/ and /-uŋ/ respectively. These morphemes are reflexes of the primary protomorpheme /-o˜/ < *-ɔw˜ with raising of the mid-vowel to /u/ triggered by the preceding high vowel of the perfective marker *-i. In RP and BH this perfective marker became null marked in the present perfective construction by [MI 56.]. Despite this loss of the trigger for raising, the high vowel /u/ remains in the BH present perfective construction. In RP, the primary first person ending has been reinstalled in the present perfective construction following the deletion of the perfective marker *-i. As a result of this change, RP does not have two primary systems—AGR.IA and AGR.IB. This reinstallation in RP is not linguistically complex, in fact it reduces the complexity of the morphological system, and thus is not diagnostic of a propagation event.

[MI 61.] > /-o˜/ '1SG' in the present perfective formation {RP} (after [MI 56.]). Supportive, not diagnostic.

The other lect with an AGR.IB (present-perfective) system is SH, which also has the higher variant /-u/. However, unlike RP and BH, SH does not have progressive raising of *o > /u/ as a phonologically general process (e.g. /d̪ekʰilo/ < *d̪ɛkʰ-il-o 'you saw'). The variation in the *first person* endings when preceded by the perfective *-i (see §6.3.5) is a morphologically-specific raising process.

[MI 62.] *-o '1SG' in AGR.I > /-u/ / iC_ {SH}. Diagnostic.

This morphologically-specific explanation could account for the higher vowel in BH also, but that option is dispreferred on the basis of simplicity of reconstruction. Progressive raising is phonologically regular in BH, and no further change is needed to account for the vowel height of BH /-uŋ/ '1SG' in AGR.IB. The arguments put forward above point to *distinct* historical processes as explanations for the higher vowel /-u/ in BH and SH AGR.IB systems.

6.4.1.2 Secondary endings

The secondary endings in KRDS are part of innovative past and future tense conjugations (see §6.2.5 and §6.2.6), and thus cannot be inherited from earlier than the proto Magadhan stage (when the new tense constructions were innovated), and may be considerably later innovations. In the AGR.IIA (past tense) systems of Table 6.18, the first person singular and plural endings are reconstructed as reflexes of proto Kamta forms *-u˜ and *-ɔ˜ respectively. Reflexes of both these forms are found in six of the eight KRDS lects—not in KS and BN—and a reflex of one of the two forms is found in KS.

These two etyma are not found with these functions in AGR.IIA (past tense) systems elsewhere in Magadhan languages that I am aware of—SCB has /-am/, Oriya has /-i/, Maithili has /-hu˜/ (possibly cognate but without Number differentiation), Bhojpuri has /-˜/, and SCA has /-o˜/ (which seems to be an extension of the primary ending rather than cognate with these distinct proto Kamta secondary endings, cf. Kakati 1962:353). In the AGR.IIB (future tense) system Bangla has /-o/ which Chatterji considers an extension of an older AGR.IIA (past tense) ending /-o˜/ in Early Middle Bangla. This ending he identifies as cognate with SCA /-o˜/, which in turn he (1926:975) and Kakati (as referenced above) derive from the primary system. This all suggests that Bangla /-o/ '1.AGR.IIB' is not cognate with *-ɔ˜ of KRDS. In addition to the unlikelihood of formal cognacy, there is also the problem of functional disjunction, given that written records attest that number was lost as a marked category of *primary endings* 'from the earliest times in Bengali' (Chatterji

1926:931), and the records give no indication that number ever was a grammaticalised category in the Bengali *secondary systems*.

Given the innovative status of the secondary endings and the etymological uniqueness to KRDS of secondary endings *-u˜ '1SG' and *-ɔ˜ '1PL', these innovative grammaticalised features are of considerable value for subgrouping. The conditioning is complex (involving cognate phonological forms in stable paradigmatic relations). It is also distinctive, given that the neighbouring lects Bangla, Asamiya and Maithili do not distinguish number in the agreement endings, nor have since 'the earliest times' (ibid.). Furthermore, there are sociohistorical conditions which can account for the original propagation of these features in a proto Kamta lect, consequently spread through migration across the present day KRDS area (cf. §7.3.1). This morphological innovation therefore is diagnostic of a propagation event:

[MI 63.] > *-u˜ '1SG', *-ɔ˜ '1PL' in AGR.IIA systems {KRDS, except BN}. Diagnostic.

This linguistically complex innovation establishes these lects to be a subgroup, as further discussed in §7.3.1. While number marking is absent in KS and BN, in the case of KS it is likely—for reasons discussed in §6.4.3—that this lect inherited number marking in agreement morphology, but has more recently generalised some of the inherited endings to include both singular and plural functions. The case of BN is less conclusive and is discussed in §6.4.3, as well as just below.

The BN ending in system AGR.IIA (past tense) is not a regular reflex of either of the reconstructed AGR.IIA forms *-u˜ and *-ɔ˜. (Neither Prosodic Vowel Raising or Progressive Vowel raising of *ɔ > /o/ occur in BN; therefore the anticipated reflexes of these reconstructed forms in BN are **-u˜, **-ɔ˜.[6]) The BN *secondary* ending /-o˜/ '1SG' appears rather to be cognate with the BN *primary* ending /-o˜/ '1SG' < *-ɔw˜. The same morpheme occurs in SCA, see further §6.4.3.

The other morpheme in the AGR.IIA (past tense) system which is noncognate is /-i/ '1PL' in SH. This is cognate with the AGR.I (present tense) ending *-i (> /-i/ in SH). The presence of this etymon in the AGR.IIA (past tense) system is the result of analogical extension:

[MI 64.] Analogical extension. *-i '1PL' in AGR.IA > /-i/ '1PL' in AGR.IIA {SH}. Diagnostic.

Within KRDS this change is unique to a contiguous subsection in the central Jalpaiguri region near SH. Based on sociohistorical plausibility it is diagnostic of a propagation event.

The AGR.IIB (future tense) endings are all cognate across the KRDS lects, barring the absence of number distinction in KS and BN. The variation across KRDS in the height of *-ɔ˜ is accounted for by regular phonological changes. Note that the nasalisation of *-ɔ˜ '1PL' is not lost in KRDS, but transferred to the tense marker on the immediate left: *bV˜ > /mV/. This is morphologically conditioned, and not phonologically general. (Compare the following examples: *ba˜ʃ 'bamboo' > /ba˜ʃ, ba˜s, baʃ/ not /maʃ/; *ba˜ʃi 'flute' > /ba˜ʃi, ba˜si, baʃi/ not /maʃi/.[7]) Changes involving the transfer of the nasal value are formulated in §6.4.1.3.

---

[6] ** indicates expected but non-occurring forms.

[7] There is an example of the exact reverse process as an irregular variation in NIA: the nasal and stop elements in Sanskrit mahiṣá 'buffalo' are separated into oral stop and nasal vowel ba˜... in several of the KRDS lects, as well as in much of NIA more generally. Compare Turner (1962–66:573, #9964).

Reproduction of **Table 6.25:** Reconstructed pKmt system of AGR endings

	AGR.I	AGR.IMP	AGR.IIA	AGR.IIB	AGR.IIC
1SG	*-ɔw̃		*-ũ̯	*-Ø < *-ũ̯	*-ũ̯
1PL	*-i			*-ɔ̃	
2SG	*-iʃ	*-ɛkɔ		*-o	
2PL	*-[ɔ;ɛ]n(t̪i)	*-ɔ		*-[ɔ;ɛ]n(t̪i)	
3	*-ɛ	*-[o;u]kɔ	*[-Ø; -ɛ]	*-ɛ	*-ɔ

The first person endings in the AGR.IIC (past definite) system are identical with those of the AGR.IIA (past indefinite) system. The AGR.IIC system is inherited from proto Kamta, but at the proto Kamta stage it differed from the AGR.IIA system only for the third person marking (see §6.4.2).

6.4.1.3 Leftwards transfer of nasal value from agreement to tense marker

The change by which the nasalisation of a vowel is transferred onto the future tense marker is summarised as follows:

[MI 65.] *-ib 'FUT' + *V̆ > /im-V/

The NIA lects with instances of such nasalisation are (to my knowledge):

- all eight of the sampled KRDS lects;
- all Asamiya varieties including SCA (Purkait 1989);
- Central, western and northern [Old] Purnia (perhaps only sporadically), loosely categorised by Grierson as transitional between Northern Maithili and Bangla (Grierson 1980 [1887]:26, 34, 36, 41);
- Varieties around Rajshahi of Bangladesh (Islam 1992; Khondakar 1998);
- Kharia Thar (but not Mal Paharia), spoken in the Rajmahal hills on the border of West Bengal and Jharkhand (D. Dasgupta 1978);
- South Dinajpur Bangla varieties (Purkait 1989);
- Varieties of north-west Midnapore (Purkait 1989);
- Some eastern Bangla varieties, namely around the Bakhar area of 'central' East Bengali, and in 'central-north' East Bengali (Haldar 1986);
- the local variety of Ramnagar police station (Purkait 1989), south from Midnapore;
- Early Oriya found in the 15th and 16th century inscriptions (Chatterji 1926:531–532);
- Modern Oriya (Chatterji 1926:532), according to Dash (1982:82) this is a 'social dialect of Cuttack';
- Magahi (Chatterji 1926:532);
- unspecified Middle Bangla dialects (Chatterji 1926:967).

These lects are spoken over quite a vast area, shown approximately by the shaded area in Figure 6.1. Note that there are other Indo-Aryan lects within the shaded area which do not share this feature.

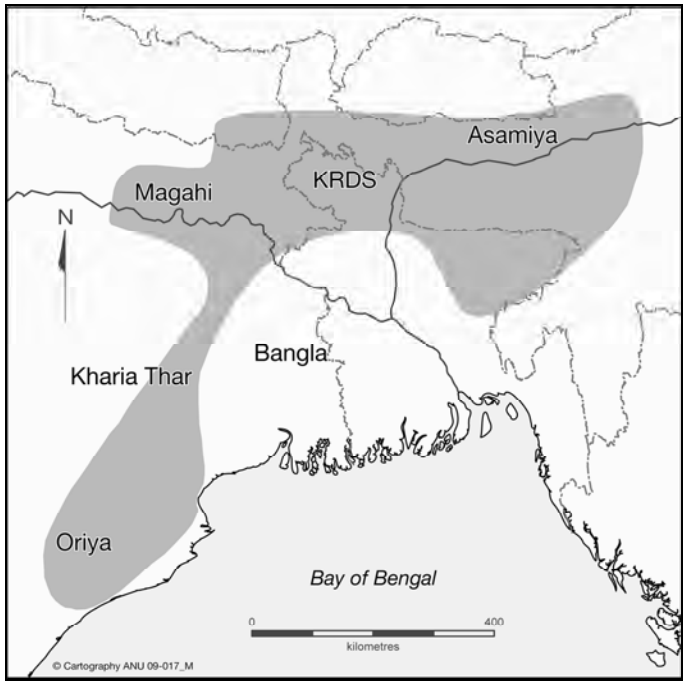


Figure 6.1: Approximate range of a nasalised future tense marker /m/

The diagnostics for propagation events are: linguistic complexity, ecological distinctiveness, and sociohistorical plausibility of propagation. The transferral of the nasal value is not linguistically complex, and furthermore is areally consistent with the eastern Magadhan tendency towards lenition of medial labial stops (cf. §4.3.5). The range is also not conducive to a sociohistorical explanation involving interconnected propagation. Therefore this change is not diagnostic of a propagation event (cf. §3.4.1). In support of this nondiagnostic judgement, it is not entirely clear from Chatterji's examples for Magahi and Oriya whether a nasalised vowel is indeed always part of the conditioning environment for nasalisation of the /-b/ future in those two lects. This does not take away from, but rather strengthens the point that the nasalisation of *-b > *-m* could have been replicated multiple times, with independent propagations. Variable nasalisation of the future tense marker *-ib may in fact have been inherited from the common Magadhan stage with independent regularisation in separate areas.

[MI 66.] *-ib 'FUT' > /-im/ 'FUT' {several Magadhan lects, see Figure 6.1}. Nondiagnostic.

Related to this nasalisation of *-ib 'future', is the fusion of secondary ending *-u˜ '1SG' with the future tense marker to give *-im:

[MI 67.] *-ib 'FUT' + *-u˜ '1SG (> *-iβ˜u, *-iw˜u > *-iw˜) > *-im 'FUT.1SG' {KRDS, south Dinajpur, SCA}. Diagnostic.

The fusion of these tense and agreement morphemes plausibly went through intermediate stages: *-ibu˜ > *-iβ˜u, *-iw˜u > *-iw˜ > /-im/. This reconstructed change process is linguistically plausible, involving coalescence of a labial vowel with a preceding labial (and lenited) stop. The conditioning of the change is relatively complex as it is restricted to first person singular, and involves fusion of categories—more complex than was the case for [MI 65.] above. Based on the data in Purkait (1989), the fusional change [MI 67.] is also considerably more limited in range than [MI 65.], and found only in the KRDS varieties, the

neighbouring South Dinajpur varieties, and the Asamiya lects. The fusional change is not found in Rajshahi varieties according to Islam (1992:143), nor in south-west Bangla varieties according to Purkait (1989). In both these cases the reported corresponding forms are *-mu* or *-mi*. Thus, in contrast to the general nasalisation change [MI 65.], the fusion of future tense and first person singular marking in [MI 67.] is more linguistically complex and found only in geographically contiguous lects. Such total fusion of tense and agreement marking is not found elsewhere in Magadhan lects that I have found. For these reasons the change [MI 67.] is diagnostic of a propagation event, while [MI 65.] is not.

A nasalisation change also occurs in the past tense morphology in a way highly similar to the nasalisation of the future tense marker by [MI 65.] (though without the fusion of tense and agreement marking). The nasal value is transferred from the agreement suffix onto the preceding past tense morpheme *-il > /-in/. The range of this change is not as widespread in KRDS as the *-ib > /-im/ change.

[MI 68.] *-il 'PST' + *-Vˇ > /-in-V/ {KS, RL, MH, TH, SH, Rajshahi, south Dinajpur, Midnapore, early Maithili, Marathi}. Nondiagnostic.

This innovative feature is also found in Rajshahi varieties (Islam 1992; Khondakar 1998), Kharia Thar (D. Dasgupta 1978), south Dinajpur, and north west Midnapore (see south west Bangla in Purkait 1989), optionally in early and modern Maithili (Jha 1985 [1958]:467), and Marathi (Masica 1991:312). The same issues of noncontiguity and noncomplexity apply as in the case of [MI 65.]. Furthermore, there is the added possibility that, given the prior nasalisation of the future tense marker, this nasalisation could have been analogically extended to the past tense marking. The multiple linguistic motivations for nasalisation of *-il make it difficult to consider it diagnostic of an interconnected propagation event. Thus the range of [MI 68.] is considered nondiagnostic.

6.4.1.4 Summary of reconstructed first person agreement endings

The first person endings and their functions in the agreement systems are reconstructed for proto Kamta as follows:

Table 6.19: Reconstructed first person agreement endings for pKmt

	AGR.I	AGR.IIA	AGR.IIB	AGR.IIC
1SG	*-ɔwˇ	*-uˇ	*-Ø < *-uˇ	*-uˇ
1PL	*-i	*-ɔˇ		

As indicated in Table 6.19 (and implied in [MI 67.]), the AGR.IIB (future) ending *-Ø is a morphologically conditioned reflex of the same etymon which gives the AGR.IIA (and AGR.IIC) ending *-uˇ '1SG' (see [MI 67.]).

6.4.2 Second and third person agreement suffixes

In this section, second and third person endings are considered side by side. The argument below is that the variation within the data of Table 6.20 is best explained by intertwined etymologies within these categories. The second person imperative forms may mark honour as much as number; hence the categories for the imperative conjugation may need revision after further synchronic study.

Table 6.20: Comparative reconstruction of KRDS second and third person agreement endings

	2nd person						3rd person				
	AGR.IMP	AGR.I		AGR.II			AGR.IMP	AGR.I	AGR.IIA	AGR.IIB	AGR.IIC
		SG	PL	SG	PL						
pKmt	*-ɛkɔ *-ɔ	*-iʃ	*-[ɛn, ɔn] [a]	*-o	*-[ɛn, ɔn] [a]		*-(o,u)kɔ	*-ɛ	*-Ø *-ɛ	*-ɛ	*-ɔ
KS	-ʌ	-is		-o				-ɛ	-ɛ	-ɛ	
RL	-ɛk -ʌ	-i	-ʌn	-o	-ʌn		-ok	-ɛ	-ɛ	-ɛ	
MH	-ɛk -ʌ -[ʌ]n	-is, -i	-ʌn	-o	-ʌn		-ok	-ɛ	-Ø, -ɛ	-ɛ	
TH	-ɛk -[ɛ]n	-i	-ɛn	-o	-ɛn		-ok	-ɛ	-Ø, -ɛ	-ɛ	o
SH	-iṭ -[ɛ]n	-iṭ	-ɛn	-iṭ [b]	-o [c]	-ɛn	-uk	-ɛ	-ɛk	-ɛ	o
RP	-ɛk -o -[ɛ]n	-iʃ	-ɛn	-u		-ɛn	-uk	-ɛ	-Ø	-ɛ	o
BH	-Ø, -ɛk -o	-iʃ	-ɛn	-u		-ɛn	-uk	-ɛ	-ɛk	-ɛ	o
BN	-[ɛn]	-is[TR], -a[INTR]		-i [d], -a [e]			-ɔk	-ɛ	-Ø	-ɔ	

[a] At the time of the proto Kamta stage this form may have been *-[ɔ,ɛ]nti, depending on the chronology of [MI 71.]. See the discussion below.

[b] Subsystem AGR.IIC.

[c] Subsystem AGR.IIB.

[d] Second person, low honour, cf. section 6.3.8.

[e] Second person, high honour, cf. section 6.3.8.

6.4.2.1 Primary endings

Reconstruction begins with the primary endings, which (as argued in the introduction to §6.4) are retentions, unlike the secondary endings. The morpheme *-iʃ '2SG' is a reflex of OIA *-asi* involving post-MIA metathesis of the /s/ and /i/ elements. The MIA form *-asi* is not the phonologically regular reflex of OIA *-asi*, which according to Bubenik would be **-ahi. He terms the MIA form a 'Prākritism' (2003:228, with Tagare 1948:288). Reflexes of the Prakritic (or semi-Tatsama) form *-asi* are found in Chattisgarhi and Marathi /-əs/, and Nepali /-es/. Chatterji also lists *-is* and *-asa* for Bhojpuri (1926:936), Tiwari has *-asi* and *-asa* (1960:168–169). As an inherited morpheme, its *position* in the primary system of agreement is a retention, but there has been an innovative metathesis. According to Chatterji (see just above), this innovation is common with Bhojpuri and thus seems to have been a variable inheritance from proto Magadhan. A reconstruction of the historical propagation of this metathesis requires broader Magadhan reconstruction than is within the scope of this study.

The morpheme *-ɛ occurs in the primary system for third person. This is the regular reflex of Sanskrit present active *-ati* '3SG' (> *-ai* > *ɛ). Cognates are found in Asamiya and Bangla /-e/, and in Maithili *-ae*, as well as in many more NIA lects. (The monophthongisation absent from Maithili is, nonetheless, widely distributed in NIA.) These are inherited morphemes in inherited morphological positions, and therefore not diagnostic of propagation events.

The last primary endings in Table 6.20 which are reconstructed for proto Kamta are *-ɛn(t̪i) and *-ɔn(t̪i), both as '2PL'. Of these, *-ɔn(t̪i) seems to be an etymological continuation of Sanskrit *-anti* '3PL', though note the change in function:

[MI 69.] *-anti* '3PL' > … > *-[ɔ;ɛ]n(t̪i)[8] '2PL' {KRDS}. Supportive, not diagnostic.

The use of /-n/ for plural marking in *third person* agreement occurs across a large portion of NIA, but it is much less common in second person plural agreement (cf. Masica 1991:263–264). The only occurrences of second person agreement involving /-n/ in Masica's data are the 2/3PL ending in Chattisgarhi, and the 3PL ending in Bangla—with a note that the same morpheme is extended to high second person. The use of /-n/ in KRDS is principally to mark *second person plural, and it is not used in third person marking*. The change in function of this inherited morpheme by [MI 69.] from third to second person plural thus seems to be distinctive of KRDS within the NIA lects. However, this shift in function of third person plural, through second person high honour, to second person plural, is not linguistically complex. Therefore [MI 69.] is listed as supportive, but not itself diagnostic of a proto Kamta propagation event.

The situation regarding second person plural marking is further complicated because while the function of the /-n/ element is uniform across KRDS, *the vowel that precedes it is not*. In TH, SH, RP and BH, the second person singular ending is /-ɛn/, but in RL and MH it is /-ʌn/. Note that the ending in {TH, SH, RP, BH} is formally similar to, though functionally distinct from, the Bangla 2/3 honorific ending /-en/. It is not out of the question that the Bangla ending has played a role in these four lects acquiring *-ɛn for second person plural. The influence of Bangla in this regard is sociohistorically plausible given that it is these same lects which reflect the influence of Bangla in other changes (cf.

[8] The *-ɛn(t̪i) variant is included in this change because of the cognacy of the *n element (see discussion that follows), and because [MI 69.] is the formulation of a change in function from third person plural to second person plural. It is this functional change which is common and distinctive to KRDS.

§7.5.3.2). However, in the case of /-ɛn/ there is an alternative explanation for its presence in the four KRDS lects, as well as Bangla, which is based on inheritance of variation, as follows.

Recall that the morphemes *-ɛn and *-ɔn are not reflexes of inherited second person endings, but have instead shifted in function from an earlier third person agreement ending *-anti*. The Apabhraṁśa third person singular ending is *-ai* > *-ɛ. A highly plausible etymological explanation for *-ɛn, therefore, is mixing of the Apabhramsa third person endings: *-ai* '3SG' and *-anti* '3PL' > *-*ainti* '3PL' > *-ɛn '3PL'. Variability between *-anti* and *-ēnti* is in fact attested in 14th century Bangla writings (Chatterji 1926:933). Therefore we can conclude that (a) mixing of inherited primary 3SG and 3PL endings is old, and (b) this mixed reflex occurred in variation with the standard reflex of OIA *-anti*, probably for quite some time. In summary: the evidence from early middle Bangla documents supports a reconstruction of the proto Kamta second person plural endings as *-[ɛn; ɔn]. The variation was part of the proto Kamta inheritance, and can be reconstructed as inherited from yet earlier protostages based on the cognates in early Middle Bangla [*-anti*; *-ēnti*]. Note also the Chattisgarhi endings [*-an*; *-en*] given by Masica (1991:264). The inheritance of variation, with subsequent regularisation, is not diagnostic of a propagation event (cf. §3.4.1.4) because of the possibility of separate, nonintegrated regularisation. However, this change may support the grouping of Bangla, Asamiya and KRDS.

[MI 70.] *-ai* '3SG' and *-anti* '3PL' > *-[*ainti,anti*] '3PL' {pre-proto Kamta and pre-proto Bangla (and plausibly pre-proto Asamiya) stage}. Supportive, not diagnostic.

The inclusion of pre-proto Asamiya in this change is discussed at the end of this section.

Regarding the change in *form* of this morpheme from early NIA *-anti, -ēnti* > KRDS *-[ɔn, ɛn], Chatterji analyses the chronology based Bangla records:

> The plural affix for the verb, «-anti», is found as «-anta, -enta» in the 15th century, and finally, by the 17th, it yields to the form «-en» (1926:133).

[MI 71.) *-nti* 'PL' in agreement endings > /-n/ 'PL' {many NIA lects}. (chronology uncertain). Diagnostic value uncertain.

The diagnostic value of this change is uncertain given that reduction of *-nt* > /-n/ in agreement endings is common across New Indo-Aryan, cf. Masica (1991:266):

> The retention of 3pl. (6) *-anti* in Oriya and of its /-n-/ element in several other languages (Punjabi, 'Lahnda', Sindhi, Kumauni, Nepali, Bengali, Kashmiri) is worth noting; in Marathi-Konkani-Sinhalese it was the *-t-* element that was retained.

It is possible that the nasal+stop cluster was still part of the plural ending during the proto Kamta stage. That hypothesis requires that the reconstructed forms be represented as *-[ɔ;ɛ]n(t̪i). No conclusion has been reached on the chronology of [MI 71.] and thus no conclusion can be given on the precise form of the ending during the proto Kamta stage. Accordingly, the curved brackets enclose material whose presence in proto Kamta is ambiguous.

Finally, some comment must be made regarding the Asamiya form /-a/ '2H' also found in the BN system. Kakati proposes that the origin of this morpheme lies in nominal rather than verbal morphology (Kakati 1962:351), but this etymology is not convincing. Nonetheless, this morpheme is present in Asamiya linguistic history at least since early

Asamiya.[9] The BN form /-a/ is diagnostic of contact relations with Asamiya, but the chronology of these contact relations—whether a recent or old borrowing—is once again ambiguous, as was seen for the genitive case in §5.3.5.

[MI 72.] > /-a/ '2H' {BN} (chronology uncertain). Diagnostic of contact relations with Asamiya.

It is possible (and perhaps likely, given the genetic relations between Bangla, Asamiya and KRDS) that Asamiya also inherited the '3PL' ending *-ɔn̪ṭi from a pre-proto Asamiya stage, and that this inherited plural marker was lost during the proto Asamiya stage at the same time as this /-a/ '2H' ending was introduced. This possibility is relevant to the range of [MI 70.].

6.4.2.2 Imperative endings

Among the imperative endings given for KRDS in Table 6.20, there are two innovative features that are diagnostic of propagation events. Firstly, there is the third person ending reconstructed as *-(o,u)kɔ. See Table 6.21 for equivalent forms in other modern Magadhan lects, as well as earlier forms from NIA literature:

Table 6.21: Third person imperative endings in Magadhan lects

	Bangla	Asamiya	Oriya	Maithili	Bhojpuri
Source	Chatterji 1926	Kakati 1962	Ray 2003, Misra 1975	R. Yadav 2003 Jha 1985 [1958]	Tiwari 1960, Verma 2003
modern form	-uk; -un	-ɒk	-u; -ɔn̪ṭu	-ɔ̃; -o; əu̪ṭʰ	-Ø; -as, -an; -i(n)
earlier form	*-u*; *åu(k)*	*-o*; *-oka*	*-ahu*	*-aü*; *-a̪ṭu*, *-ao*, *-̪tu*; *-a*, etc.	

The /k/ element of this third person imperative suffix is pleonastic, innovative, and unique to KRDS, Bangla and Asamiya out of the Magadhan lects:

[MI 73.] third person imperative ending suffixed by + *-kɔ > *-Vkɔ '3.IMP' {SCB, SCA, KRDS}. Diagnostic.

The precise qualities of the vowel to be reconstructed are obscure, with /u/ found in SCB and some KRDS lects, /ɒ/ in SCA, /ɔ/ in BN, and /o/ elsewhere in KRDS. This variation in the vowel is probably not the reflex of a unique third person ending. Nevertheless the addition of the pleonastic *-kɔ is firmly attested for all these lects. Note that the suffix is reconstructed with a final *-ɔ, which accords with the written records, and whose loss is expected by general NIA phonological changes to final vowels (cf. §4.4.11). The innovation [MI 73.] introduces a new segment to the inherited morpheme, which entails a certain linguistic complexity of conditioning and is diagnostic of a propagation event involving Bangla, Asamiya and KRDS.

---

[9] The presence of /-a/ 'second person' in Maharaja Nara Narayana's letter of 1555 AD (cf. §7.3.1.3.) suggests this morpheme may be a proto Kamrupa inheritance which was subsequently lost during proto Kamta or post-proto Kamta.

The imperative ending *-ɛkɔ ‘2SG’ is innovative and unique to KRDS along with Hajong (according to Chatterji 1926:990). It is not reported elsewhere in eastern Magadhan varieties (that I have been able to find), including the neighbouring Rajshahi lect according to Islam (1992). The equivalent morphemes in other Magadhan lects are as follows:

Table 6.22: Second person singular imperative endings in Magadhan lects

	Bangla	Asamiya	Oriya	Maithili	Bhojpuri
Modern	-Ø, -o, -un (H)	-Ø, -a (H)	-Ø, -ɔ	-əh, -ə, -u, -o, -Ø	-e, -ə, -u
Earlier	*-a, -aha,* *-ā* (H)	*-a, -sa,* *-ā, āhā* (H)	*-a, -aha*	*-aha, -eha, -a, -hu, -ai,* etc.	

Based on the ecological distinctiveness of this use of the pleonastic /-k/ < *-kɔ in second person imperatives, as well as linguistic complexity, the following innovation is diagnostic of a propagation event.

[MI 74.] In AGR.IMP, *-ɛ ‘2SG’ + *-kɔ > *-ɛkɔ ‘2SG’ {KRDS, some Hajong lects}. Diagnostic.

The other second person imperative ending is *-ɔ ‘2PL’, which seems to be etymologically related (if indirectly) to the secondary ending *-o ‘2SG’. The argument for this etymology is somewhat complicated, and revolves around similarities between KRDS and Oriya. The imperative ending *-ɔ ‘2PL’ is cognate with Bangla imperative /-o/ ‘2NT’ (with prosodic vowel raising in Bangla, cf. §4.4.4), Oriya /-ɔ/ ‘2PL’ (cf. Misra 1975:151–152), Maithili and Bhojpuri /-ə/ ‘2NT’. Chatterji derives these forms from OIA indicative *-atha* through Apabhraṁśa *-aha* (which varies in MIA with the *-ahu* form given in Table 6.17, cf. Chatterji 1926:905–906). KRDS *-ɔ ‘2PL’ (along with its Magadhan cognates) has entered the imperative system by extension from the primary system. The presence of this analogical extension (primary > imperative) across Magadhan languages and throughout their recorded histories proves this extension to be a Magadhan or pre-Magadhan change and thus not relevant to post-Magadhan subgrouping. The etymological complexity emerges because the same etymon from the OIA primary system, *-atha,* (proposed above to give Oriya imperative /-ɔ/ ‘2PL’) is also proposed as the source of Oriya /-u/ ‘2SG’ in the secondary system. This problem is partially resolved in §6.4.2.3.

Finally for the imperative endings, the use of *-(ɛ,ɔ)n in imperative function seems to be an innovative extension of the primary endings to the imperative system.

[MI 75.] *-(ɛ,ɔ)n ‘2PL’ in AGR.I > *-(ɛ,ɔ)n ‘2PL’ in AGR.IMP {KRDS, …}. Nondiagnostic.

This analogical extension is linguistically natural and noncomplex, and not diagnostic of a propagation event.

6.4.2.3 Secondary endings

The third person endings in the secondary system preserve traces of an erstwhile transitivity distinction. Chatterji (1926:93) provides the Magadhan perspective: ‘differentiation between transitive and intransitive verbs, 3 person only … can be called a common Magadhan trait, having its germ in the Māgadhī Apabhraṁśa.’

The transitivity distinction in agreement marking is generally not retained in KRDS. However, there is a trace of this old distinction in the differentiation of third person marking between the AGR.IIA (simple past) system and the AGR.IIC (past definite) system. The latter is based on an old compound formation with auxiliary verb *aʧʰ- 'be present' and takes the erstwhile intransitive third person suffix *-ɔ.

As for the plural ending, the analogical extension of the reflexes of Apabhramsa *-anti* from the primary endings to the secondary systems is not unique to KRDS, but also occurs in Oriya /-ɔnṭi/, Maithili and Magahi /-nh-/, Bhojpuri /-ən/ and Bangla /-en/.

[MI 76.] *-anti* '3PL' in AGR.I > '3PL' in AGR.IIA and AGR.IIB {KRDS, Bangla, Oriya, Maithili, Magahi, Bhojpuri, …} (before change in function by [MI 69.]). Nondiagnostic.

It is not clear what change events should be reconstructed to make sense of this distribution which is scattered across Magadhan lects. Chatterji holds that during the early stage of 'the neo-Magadhan speeches', secondary affixation 'was not indispensable' (ibid. 971), that is, it was variable. The most plausible explanation seems to be that the extension of *-anti* '3PL' to the secondary systems had begun as a variable change early in the post-Magadhan period. This variation was inherited into the Magadhan languages during the period when secondary systems were variable and 'not indispensable', and then independently regularised.

Reproduction of **Table 6.25:** Reconstructed pKmt system of AGR endings

	AGR.I	AGR.IMP	AGR.IIA	AGR.IIB	AGR.IIC
1SG	*-ɔw̃		*-ũ̯	*-Ø < *-ũ̯	*-ũ̯
1PL	*-i			*-ɔ̃	
2SG	*-iʃ	*-ɛkɔ		*-o	
2PL	*-[ɔ;ɛ]n(ṭi)	*-ɔ		*-[ɔ;ɛ]n(ṭi)	
3	*-ɛ	*-[o;u]kɔ	*[-Ø; -ɛ]	*-ɛ	*-ɔ

The last secondary etymon to be examined is *-o '2SG'. The etymology of this morpheme is somewhat more difficult as foreshadowed above in §6.4.2.2. Asamiya and Bangla have noncognate forms for this secondary category: SCA /-i/, SCB /-iʃ/ and /-i/. Oriya has /-u/ across primary and secondary systems, with etymology reconstructed by Misra (1975:135) as follows:

> 2nd *-u* e.g. *khāu* < OIA *khādathaḥ* (OIA dual > plural in MIA which was transferred to singular in Oriya) > *khāaho* > *khāho* > *khāhu* > *khāu*

The presence of /-u/ in Oriya secondary systems is then by analogical extension from the primary system (ibid.:140, 143).

Comparing Misra's etymological hypotheses for the secondary Oriya ending /-u/ '2SG' and the Oriya imperative ending /-ɔ/ '2PL', she has reconstructed a common etymology as reflexes of OIA *-atha* > *-aha* for these two distinct morphemes, see Table 6.23.

Table 6.23: Etymology of second person singular endings in Oriya and KRDS

Prakrit	*-aha* ‘2PL’	*-aha* ‘2PL’
Etymological progression:	source of Oriya *imperative* ending by: regular phonological reflex *aha* > *a* > ɔ	source of Oriya *primary* and *secondary* endings by: MIA morphological change *aha* > *ahu*
Oriya	/-ɔ/ ‘2PL:IMP’	/-u/ ‘2SG:AGR.II’
cognate with: pKmt	*-ɔ ‘2SG:IMP’	*-o ‘2SG:AGR.II’

The imperative ending in Oriya (middle column in Table 6.23) is, according to Misra, the regular reflex of the Prakrit form. The secondary ending /-u/ (extended from the primary ending) is apparently the reflex of a morphologically conditioned change *aha* > *-ahu* that occurred during MIA (see Bubenik 2003:227–228). This morphological change results in the Apabhraṁśa form *-ahu* ‘2PL’ (see Table 6.17). It is plausible that the KRDS forms have the same etymologies as the Oriya forms, given that in KRDS there is also a difference in height between the second person singular imperative and secondary endings. The historical veracity and chronology of this hypothesis, which involves different reflexes of OIA *-atha* regularised in different morphological positions during MIA, should be tested against the MIA records. Such testing is left for further research.

What remains very much part of this study is to consider the implications of this hypothesis for KRDS-Oriya historical relations. Given that the secondary systems are late- and post-Magadhan innovations, the analogical extension of *-ahu* to the secondary system is innovative, and its diagnostic value must be considered. The innovation also involves a change in the function of the inherited MIA ending.

[MI 77.] *-ahu* ‘2PL’ in AGR.I {late MIA} analogically extended to the secondary system to give *-o ‘2SG’ {proto Kamta} and /-u/ ‘2SG’ {Oriya}. Nondiagnostic.

The threefold test for diagnosing propagation events is linguistic complexity, ecological distinctiveness, and geographical contiguity or other sociohistorical explanation for range of propagation. The most significant factor in this case is the geographical non-contiguity of Oriya and KRDS. There are no clear sociohistorical events which suggest significant interaction between these two historical kingdoms, geographically separated by modern Bengal and the earlier kingdom of Gaur. The innovation [MI 77.] would seem therefore either to be a case of independent innovation and propagation on the part of Oriya and KRDS, or alternatively, a proto eastern Magadhan innovation retained in Oriya and KRDS, but lost in Asamiya and Bangla. However, retention from proto eastern Magadhan is fairly implausible given the late origin of these secondary systems (cf. §6.4 above). In summary, the proto Kamta morpheme *-o ‘2SG’ in the secondary system is plausibly cognate with Oriya /-u/ ‘2SG’, but the extension to the secondary system does not appear to be diagnostic of a common propagation event in the linguistic histories of Oriya and KRDS.

Note that this reconstruction of the development of KRDS *-o differs slightly from Chatterji’s hypothesis of a direct connection between ‘North Central’ /-u/ and Oriya /-u/ (cf. e.g. 1926:980). While supporting his hypothesis that the forms are cognate, this reconstruction argues that in the case of KRDS lects with /-u/ ‘2SG’ (e.g. RP and BH), this

form is diachronically more closely related to other KRDS lects which have /-o/ (e.g. MH and RL). Therefore the proto Kamta morpheme is reconstructed as *-o '2SG'. The high vowel /-u/ in central KRDS is the result of progressive vowel raising (cf. §4.4.2).

6.4.2.4 Summary of reconstructed 2 and 3 endings

The second and third person agreement endings are reconstructed for proto Kamta as shown in Table 6.24:

Table 6.24: Reconstructed second and third person agreement endings

	AGR.I	AGR.IMP	AGR.IIA	AGR.IIB	AGR.IIC
2SG	*-iʃ	*-ɛkɔ	*-o		
2PL	*-[ɔ;ɛ]n(t̪i)	*-ɔ	*-[ɔ;ɛ]n(t̪i)		
3	*-ɛ	*-[o;u]kɔ	*[-Ø; -ɛ]	*-ɛ	*-ɔ

6.4.3 Reconstructed proto Kamta agreement systems

Sections 6.4.1–6.4.2 present the argument that the agreement systems of the KRDS lects (with the possible exception of BN) are reflexes of a single historical agreement system (termed 'proto Kamta' and dated in §7.3.1). The proto Kamta agreement system is summarised in Table 6.25.

Table 6.25: Reconstructed pKmt system of AGR endings

	AGR.I	AGR.IMP	AGR.IIA	AGR.IIB	AGR.IIC
1SG	*-ɔw̃		*-ũ̯	*-Ø < *-ũ̯	*-ũ̯
1PL	*-i			*-ɔ̃	
2SG	*-iʃ	*-ɛkɔ		*-o	
2PL	*-[ɔ;ɛ]n(t̪i)	*-ɔ		*-[ɔ;ɛ]n(t̪i)	
3	*-ɛ	*[-o;u]kɔ	*[-Ø; -ɛ]	*-ɛ	*-ɔ

This reconstructed proto Kamta system is now compared with the contemporary systems of KS and BN which are most divergent from it, and whose status as direct descendents is consequently in doubt.

KS lacks number marking, and is thus a simpler system than reconstructed for proto Kamta. However, all of the affixes in the KS system (Table 6.8) are retentions from the reconstructed proto Kamta system. The forms retained in KS are a mix of proto Kamta singular and plural forms, and not just the singular forms, or just the plural forms. Most notably for subgrouping purposes, the KS system includes: (1) the fused future tense marker /-im/ '1SG' which proves its inclusion in the KRDS-Asamiya subgroup established by [MI 67]; and also (2) the secondary ending /-o/ '2SG' which is consistent with the independent KRDS and Oriya innovations formulated by [MI 77.]. Therefore, the KS agreement system is perspicuously reconstructed as a direct descendant of the proto Kamta agreement system. The following changes account for the divergence of KS:

[MI 78.] Loss of Number. AGR.I/AGR.IMP *-i ‘1PL’ > /-i/ ‘1’.

[MI 79.] Loss of Number. AGR.IMP *-ɔ ‘2PL’ > /-ʌ/ ‘2’.

[MI 80.] Loss of Number. AGR.I *-iʃ ‘2SG’ > /-is/ ‘2’.

[MI 81.] Loss of Number. AGR.IIA *-u˜ ‘1SG’ > /-u/ ‘1’.

[MI 82.] Loss of Number. AGR.IIA/IIB *-o ‘2SG’ > /-o/ ‘2’.

[MI 83.] Loss of Number. AGR.IIB *-Ø ‘1SG’ > /-Ø/ ‘1’.

This loss of number marking may be partially linked to language contact with the Bihari lects, as they lack number marking in first person agreement endings. However, this does not explain the change in second person marking.

BN is also considerably divergent from the proto Kamta agreement system. In the case of BN though, the system contains affixes not present elsewhere in KRDS and thus not reconstructed as part of the proto Kamta inheritance:

- /-a/ ‘2.H’ in primary and secondary systems, compared with proto Kamta *-(ɛ;ɔ)n ‘2PL’;[10]
- /-i/ ‘2.L’ in secondary systems, compared with proto Kamta *-o ‘2SG’;
- /-ɔk/ ‘3’ in the imperative, which is identical with SCA, but slightly different to other KRDS lects which have /-ok/ or /-uk/.

The BN agreement system is, with the exception of the primary morpheme /-is/ ‘2.L’, identical to the SCA system. The presence in BN’s secondary system of endings cognate with Asamiya cannot be a feature retained from Magadhan, because the secondary conjugations are post-Magadhan innovations. This leads to a confused phylogeny: BN shares some diagnostic proto Kamta morphological changes (cf. [MI 31.]), but its verbal morphology is almost identical with SCA. Phonological changes also showed a clear phylogenetic relation between BN (the ‘Koch Rajbanshi’ lect of Bongaigaon) and the rest of Asamiya. These results indicate a mixed linguistic history for BN, involving relations both with the proto Kamta stage as well as Asamiya linguistic history. The special case of BN within the eight sample KRDS lects will be returned to in the next chapter—in the context of reconstructing the sociohistorical conditioning of propagation events in linguistic history (see §7.5.4.2).

The innovations that have been reconstructed in this chapter as diagnosing PEs, or supporting the diagnosis of PEs, are as follows:

[MI 47.] *-i ‘PFV’ in simple verbs > /-i/ ‘PFV’ in both simple and compound verb constructions {BH, RP, SH} (before [MI 56.]). Diagnostic.

[MI 50.] *-iba ‘INF’ replaced with /-na/ ‘INF’. Diagnostic of contact relations through diglossia with Hindi.

[MI 51.] *-iba ‘INF’ + *-[ɛ]r ‘GEN’ > *-ibar ‘INF’{SH, RP, BH}. Diagnostic.

[MI 52.] *-ibar ‘INF’ > /-ir/ / C_ {BH}. Diagnostic.

[MI 53.] *-ʧʰ ‘PRS’ (>*-s) > /-ʃ/ in present and past perfective formations {SH}. Diagnostic.

[10] But compare §7.3.1.3.

[MI 54.] *VERB-i-ʧʰ-AGR.I ‘present perfective’ replaced by VERB-e-ʧʰ-AGR.I{TH}. Diagnostic of contact relations with SCB.

[MI 55.] *VERB-ɛ-ʧʰ-AGR.I ‘present continuous’ replaced by VERB-ʧʰ-AGR.I {TH}. Diagnostic of contact relations with SCB.

[MI 56.] *VERB-i-s-AGR.I ‘present perfect’ > VERB^{H}-s-AGR.I {RP} and VERB^{H}-s-AGR.IA {BH} (after [MI 47.] and [PI 33.]). Diagnostic.

[MI 57.] VERB-INF + present perfective of *lag- ‘attach’ > ‘present continuous’ {RP, BH, BN}. Supportive, not diagnostic.

[MI 58.] > VERB-ʧʰi[l;n]-AGR.IIC ‘present continuous’ {TH}. Diagnostic of contact relations with SCB.

[MI 59.] > VERB-PFV *aʧʰ-il-AGR.IIA ‘past continuous’ {BN, BH}. Diagnostic value ambiguous between contact relations with Asamiya or a PE within BH and BN.

[MI 60.] Introduction of patient agreement in verbal morphology {RL}. Diagnostic.

[MI 61.] > /-o˜/ ‘1SG’ in the present perfective formation {RP} (after [MI 56.]). Supportive, not diagnostic.

[MI 62.] *-o ‘1SG’ in AGR.I > /-u/ / iC_ {SH}. Diagnostic.

[MI 63.] > *-u˜ ‘1SG’, *-ɔ˜ ‘1PL’ in AGR.IIA systems {KRDS, except BN}. Diagnostic.

[MI 64.] Analogical extension. *-i ‘1PL’ in AGR.IA > /-i/ ‘1PL’ in AGR.IIA {SH}. Diagnostic.

[MI 67.] *-ib ‘FUT’ + *-u˜ ‘1SG (> *-iβ˜u, *-iw˜u > *-iw˜) > *-im ‘FUT.1SG’ {KRDS, south Dinajpur, SCA}. Diagnostic.

[MI 69.] *-anti* ‘3PL’ > … > *-[ɔ;ɛ]n(t̪i) ‘2PL’ {KRDS}. Supportive, not diagnostic

[MI 70.] *-ai* ‘3SG’ and *-anti* ‘3PL’> *-[*ainti,anti*] ‘3PL’ {pre-proto Kamta and pre-proto Bangla (and plausibly pre-proto Asamiya) stage}. Supportive, not diagnostic..

[MI 72.] > /-a/ ‘2H’ {BN} (chronology uncertain). Diagnostic of contact relations with Asamiya.

[MI 73.] Third person imperative ending suffixed by + *-kɔ > *-Vkɔ ‘3.IMP’ {SCB, SCA, KRDS}. Diagnostic.

[MI 74.] In AGR.IMP, *-ɛ ‘2SG’ + *-kɔ > *-ɛkɔ ‘2SG’ {KRDS, some Hajong lects}. Diagnostic.

7 *Historical sociolinguistic reconstruction*

7.1 Introduction

This final chapter of historical reconstruction brings together the threads from Chapters 4–6 within the framework of historical sociolinguistic reconstruction, culminating in a coherent account of KRDS linguistic history. The sociohistorical framework for establishing and sequencing PEs (proceduralised in Chapter 3) involves six steps—three of which have been completed in Chapters 4–6, and three of which remain to be worked out in this chapter. The tasks remaining for this chapter are those in italics:

I. Reconstruct the directionality of linguistic changes (e.g. by the CM).

II. *Scrutinise in as much detail as possible the social and geographical ranges of the linguistic innovations established under Step I.*

III. Apply the three diagnostics (linguistic complexity, ecological distinctiveness, and sociohistorical plausibility) to the innovations reconstructed under Step I to establish Propagation Events in linguistic history.

IV. Investigate whether the chronology of any PEs that result from Step III can be established (a) by linguistic seriation involving *necessary* diachronic dependency or *plausible* diachronic dependency (cf. §3.4.3.1), or (b) by textual sequencing.

V. *Consider (i) the possible permutations of SCEs (divisions and integrations) which would account for the disjunction in PNs, (ii) the relative sociohistorical plausibility of each possible permutation, and (iii) the relative sociohistorical plausibility of a SCE as against the co-existence of the PNs within a complex SC. Accordingly, reconstruct the chronology of PEs by selecting the most plausible sociohistorical explanation.*

VI. *Use the chronologies established by sociohistorical linguistic seriation (Step V), as well as linguistic seriation and textual sequencing (Step IV) to reconstruct an account of the linguistic history.*

Step II involves dialectology, and Step V the reconstruction of SCEs. These two steps utilise distinct sets of data and principles of analysis but are jointly presented here in the context of Step VI: a coherent account of linguistic history resulting from linguistic and sociohistorical methods of reconstruction.

This chapter, given its subject matter, has the potential to be quite open ended. The discipline of sociolinguistics testifies to the complex social conditioning of the propagation of linguistic variants. The dialectological data outlined in this chapter are intentionally limited: limited to the *deshi* 'localised' end of the social domain, limited to a certain number of speakers, and limited to a certain number of tokens for each variant. This renders a conservative dialect geography which leaves room for further dialectological and sociolinguistic studies, but is sufficient for our purposes here. Limitations not withstanding, the data presented here and in the appendices constitute the most detailed and systematic dialectological work undertaken of the area to date.[1]

The history that emerges in this chapter is of the formation of the proto Kamta speech community and its language—ancestral to all contemporary KRDS lects—as the linguistic neighbour of the proto Gauḍa (Bangla) and proto Asamiya SCs and languages. The Propagation-Defined Languages (cf. §3.2) spoken by these three proto SCs were the phylogenetic descendants of a common stage which I argue below (§7.3.3) is yet to be properly established by historical reconstruction. The emergence of the proto Kamta SC was followed by internal divisions, as well as integrations with neighbouring SCs, and resulted in the contemporary lects (language[s] or dialects) known as Kamta, Rajbanshi and Deshi Bhasha. The history of KRDS has been divided into three stages: old Kamta (reconstructed in §7.3), middle KRDS (§7.4), and modern KRDS (§7.5), The full 'tree' diagram of this linguistic history is laid out in §7.5 (schematically adjusted as per §3.4.4).

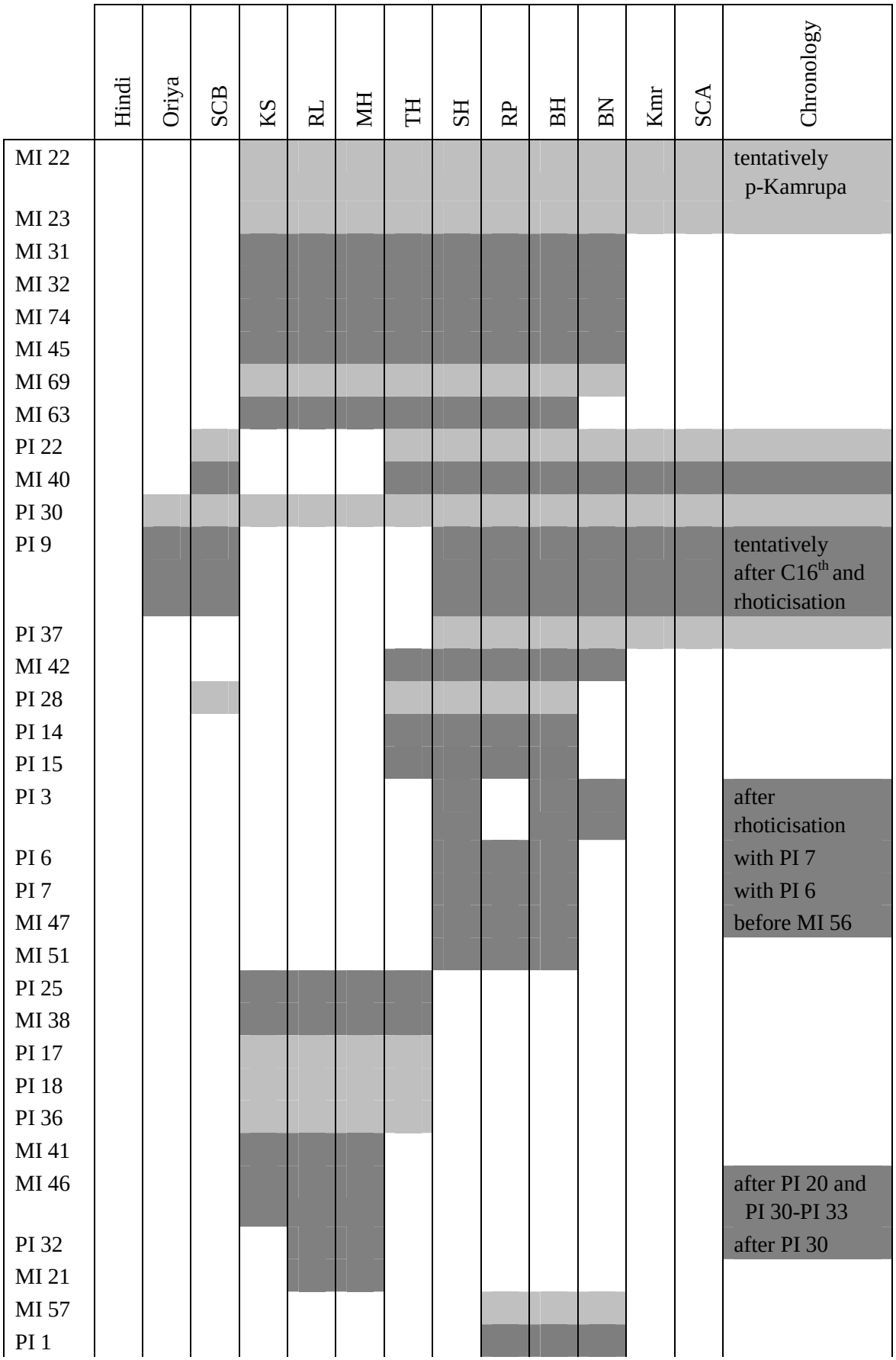
7.2 Reconstructed changes diagnostic of PEs (and SCEs)

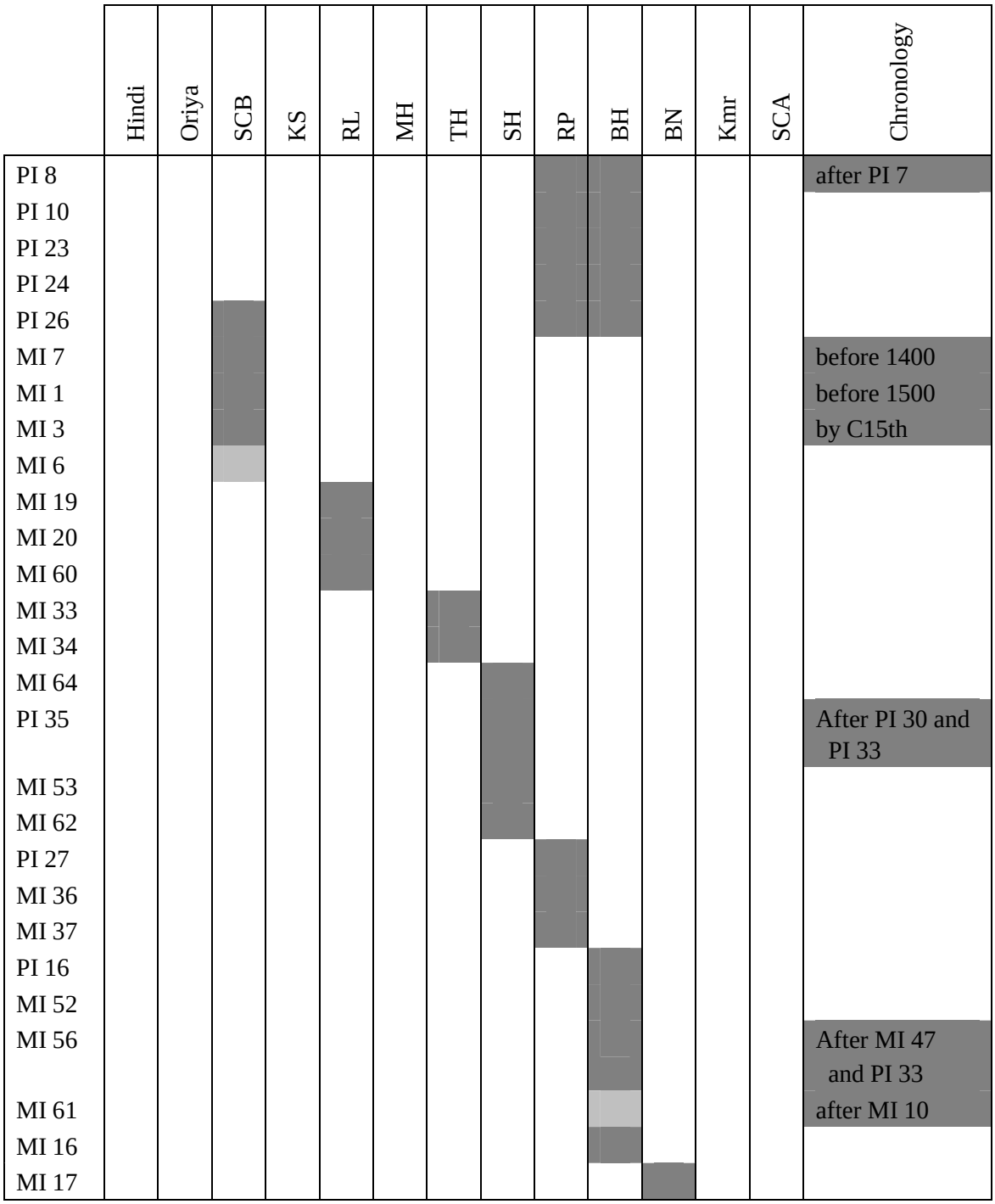
The reconstructed changes which are either (i) diagnostic of a Propagation Event and its Propagation Network, or (ii) supportive of PEs diagnosed by other changes, are shown in a tableau below. The changes shown do not include those that result from diglossia, which are instead given separately in §7.5. Dark shaded cells in Table 7.1 indicate a diagnostic change, while light shaded cells indicate a supportive but non-diagnostic change. The changes have been sorted to bring together rows with a similar range. The arrangement of rows in Table 7.1 is not indicative of sequencing.

Table 7.1: Tableau of unsequenced Propagation Networks

	Hindi	Oriya	SCB	KS	RL	MH	TH	SH	RP	BH	BN	Kmr	SCA	Chronology
PI 12														C15th or later
PI 38														uncertain
MI 73														
MI 2														
MI 70														
MI 67														

[1] Discussion of the selection of test sites, test items and subjects is in Chapter 2.


Hindi
Oriya
SCB
KS
RL
MH
TH
SH
RP
BH
BN
Kmr
SCA
Chronology
MI 22
MI 23
MI 31
MI 32
MI 74
MI 45
MI 69
MI 63
PI 22
MI 40
PI 30
PI 9
PI 37
MI 42
PI 28
PI 14
PI 15
PI 3
PI 6
PI 7
MI 47
MI 51
PI 25
MI 38
PI 17
PI 18
PI 36
MI 41
MI 46
PI 32
MI 21
MI 57
PI 1
tentatively p-Kamrupa
tentatively after C16th and rhoticisation
after rhoticisation
with PI 7
with PI 6
before MI 56
after PI 20 and PI 30-PI 33
after PI 30



The changes in Table 7.1 are now re-ordered in a simplified tableau to give some idea of the 'shape' of the linguistic phylogeny which will be argued for in the rest of this chapter. The lefthand column shows the reconstructed chronological stages of KRDS linguistic history. The chronological and genetic relations between lects are unclear using this method of portrayal, and more exact genetic relationships are displayed in a tree diagram (Figure 7.20). Innovations found in only one KRDS lect are excluded from Table 7.2 for simplicity of presentation.

Table 7.2: Propagation networks in (roughly) sequential order

		Hindi	Oriya	SCB	KS	RL	MH	TH	SH	RP	BH	BN	Kmr	SCA
pre-AD 1250	MI 73 MI 2 MI 70 MI 67 MI 22 MI 23													
1250 – 1550	MI 31 MI 32 MI 74 MI 45 MI 69 MI 63 MI 7 MI 1 MI 3 MI 6													
1550 – 1787	PI 6 PI 7 MI 47 MI 51 PI 25 MI 38 PI 18 PI 36 PI 37 PI 8 PI 23 PI 24 PI 26 PI 10 PI 9 PI 30 PI 12 PI 38													
1787 – present	PI 32 MI 21 MI 41 MI 46 PI 17 PI 28 PI 14 PI 15 MI 42 PI 22 MI 40 MI 57 PI 1													

The remainder of this chapter presents a sociohistorical argument in support of this chronology of linguistic innovations. There are several further changes, diagnostic of contact relations rather than PEs, which are examined in §7.5.

7.3 The origins of the 'old Kamta' speech community, and its relation to early Bangla and Asamiya: AD 1250–1550

Several views have been put forward previously on the origin of the KRDS language and dialects, but none have been supported by in-depth, systematic and detailed reconstruction of KRDS linguistic history.[2] Furthermore, purportedly historical classifications have not always been based on robust historical methodology such as the principles of (i) classifying on the basis of innovations and not retentions, and (ii) excluding innovations that are plausibly the result of independent and parallel propagation events (cf. §3.4). For example, Grierson held that Rajbangsi (an alternative Romanisation of রাজবংশী) was an *eastern* variety of Bangla (1903–28:vol.4, p.18). However, the areas classed by Grierson as north and east Bengal have similar linguistic ecologies, marked by contact with Tibeto-Burman languages. Therefore, there is a strong possibility that certain similarities between the Indo-Aryan of north and east Bengal are the result not of unified PEs, but of independent and parallel changes. The structural similarities between the changes are the result not of propagation, but of similar starting conditions and similar linguistic ecologies (cf. §3.4.1.2). As a result, Grierson's statement of relatedness between KRDS and eastern Bangla is not historically robust.

The received wisdom on KRDS's historical origins and relations boils down to two propositions.[3] Firstly, that an Indo-Aryan language spread from *Māgadha* (an area of today's Bihar) eastwards and northwards to *Kāmatā* (now pronounced [kaṃta], that is today's North Bengal) and from there further east to *Kāmarupa* (that is today's Assam). The implication of this position, which seems to have been initiated by Grierson, is that KRDS and Asamiya share a common linguistic ancestor and thus constitute a subgroup. The second piece of received wisdom is that Bangla and Asamiya (and KRDS) constitute a linguistic subgroup (Bengali-Assamese), and are more closely related to each other than to Oriya. This claim regarding subgrouping within the eastern Magadhan languages was advocated by Chatterji and others who followed him. Both propositions are critically examined below in the light of this reconstruction of KRDS linguistic history.

7.3.1 The proto Kamta speech community: approximately AD 13th–16th centuries

The first question to be addressed, as a keystone for understanding KRDS linguistic history, is the sequencing of (a) those PEs which uniquely define the KRDS lects, and (b) those which define smaller divisions within KRDS.

There are six changes identified by this study which uniquely define KRDS. All of these changes are morphological, and five are diagnostic of PEs:

---

[2] The great linguistic studies that have made claims regarding KRDS (especially Chatterji 1926) are without question in-depth and systematic, but their subject matter is not KRDS linguistic history per se. While they treat their own subject with rigour, they make no special study of KRDS, and treat it only peripherally and in passing.

[3] The broader and older historical relations—such as between proto-eastern Magadhan and other Magadhan lects, or the historical relation to still earlier stages of Indo Aryan—fall outside the scope of this study.

[MI 31.] *m^{ɦ}ra reinterpreted as ‘PL.NOM’ in pronoun system, and extended as such to third person *[o;u]mra {KRDS, also some Hajong lects}. Diagnostic.

[MI 32.] *m^{ɦ}a- reinterpreted as ‘PL.OBL’ in pronoun system, and extended to third person *[o;u]m^{ɦ}a- {KRDS, also some Hajong lects}. Diagnostic.

[MI 45.] *-ʈ^{h}ɛ + kuna > *-ʈ^{h}ɛkuna ‘place’ as a base of locational pronominals {KRDS}. Diagnostic.

[MI 63.] > *-u˜ ‘1SG’, *-ɔ˜ ‘1PL’ in AGR.IIA systems {KRDS, except BN}. Diagnostic.

[MI 69.] *-anti* ‘3PL’ > … > *-[ɔ;ɛ]n(t̪i) ‘2PL’ {KRDS}. Supportive, not diagnostic

[MI 74.] In AGR.IMP, *-ɛ ‘2SG’ + *-kɔ > *-ɛkɔ ‘2SG’ {KRDS, some Hajong lects}. Diagnostic.

While the PEs indicated by these innovations are diagnostic of a PN (and Propagation-Defined Language), there are also PEs which have more limited ranges (cf. §7.4–§7.5). For example, the palatalisation PE is diagnostic of a central KRDS Propagation Network and Language.

The disjunction between the entire-KRDS PN and more localised PNs may have resulted from several theoretically-possible sociohistorical scenarios. Firstly, the disjunction between PNs could reflect different levels of inclusiveness within a complex SC. Secondly, the disjunction could reflect earlier SC unity followed by SC division. Thirdly, the disjunction could reflect earlier SC differentiation followed by SC integration. Fourthly, the disjunction between PNs could reflect a combination of the first three scenarios. Each different sociohistorical explanation has different implications for the sequencing of linguistic history. These alternative explanations are evaluated below in the light of the known social history of the area.

Alongside considerations of sociohistorical plausibility, the sociohistorical reconstruction is also informed by (though not *determined* by) the principle of economy: *reconstruct the fewest number of SCEs necessary to account for the disjunctions between PNs*. This means that multiple PEs with the same range of propagation will be treated as diagnostic of the same PN unless there are textual or linguistic reasons to discount such a reconstruction. Therefore, we start with the hypothesis that the six KRDS-wide PEs listed above diagnose a single PN and its Propagation-Defined Language—termed ‘proto Kamta’ for sociohistorical reasons given below. (Recall from §3.4 that each Propagation Event diagnoses a Propagation Network of speaker interaction. The utterances used by this PN are termed the Propagation-Defined Language. These are the key concepts used to reconstruct linguistic phylogeny).

The KRDS-wide PN reflects:

- a period of social integration that is phylogenetically ancestral to the present-day KRDS lects, but not to other eastern Magadan lects; and
- a period of social division between this proto Kamta (KRDS) speech community and the neighbouring proto Bangla and proto Asamiya SCs.

There are two key sociohistorical events which, when taken together, put bookends on either side of a period of history that plausibly gave rise to the proto Kamta SC: *Aryanisation* and *kingdom expansion*.

7.3.1.1 *Terminus post quem*: Aryanisation of the Kamta region

Regarding the first of these historical events, it is generally accepted that the ancestors of the KRDS community were speakers of Tibeto-Burman languages (such as continue to be spoken by the Boro and Rabha peoples of the same area).[4] Aryanisation of language was accompanied by Aryanisation of culture, though the temporal coincidence of the two was not necessarily exact. It is important to distinguish between the course of Aryanisation among the *ruling classes* and that among the *peasant classes*, with the Aryanisation of the former not necessarily entailing immediate Aryanisation of the latter. Archaeological evidence suggests that Aryanism was established quite early in the *Pundra-Vardhana* (or *Barendra*) region, which is today's north-central Bengal (south of where KRDS is spoken).

> It will appear that Aryanism spread in Bengal first in Puṇḍra (North Bengal) and next through that country to Vanga (South and South-East Bengal). That the above areas of Bengal already became strongholds of Aryan culture in the third and second centuries BC is also suggested by epigraphic evidence. The earliest epigraphic record so far discovered in Bengal comes from Mahasthan (ancient Puṇḍranagara in the present Bogra district of North Bengal in East Pakistan). The inscription is written in the Prakrit language and in the Brāhmī script assignable to the second century BC. The popularity of Prakrit exhibited by the above inscriptions no doubt points to the considerably strong hold of Aryanism in the northern and eastern parts of Bengal in the centuries before Christ. (Sircar 1952:172, cited in Clarke 1969:179)

The Pundra-Vardhana region was geographically adjacent to, but did not include, the present KRDS-speaking area—which instead fell within the kingdom of Prāgjyotisha, later known as Kāmrupa. The border between these two polities generally corresponded to the course of the Karatoya-Tista river system (Clark 1969:191; see also R.C. Majumdar and J.N. Sarkar 1943:24–25). While Buddhist rulers reigned in Pundra-Vardhana, many Brahmins sought refuge in Kamrupa with its Hindu rulers (Clark 1969:180; Choudhury 1966:423). Barua reports this immigration of Brahmins to have continued 'right up to the Ahom period', that is around AD 1200 (Barua 2003:142). However, the extent of Aryanisation in Kāmrupa seems to have been limited to the ruling classes, and excluded the general populace:

> Although it is evident that the ruling houses of ancient Assam took on the framework of Hinduism very early, it is far more difficult to decide the speed with which the more ordinary people of Assam and North Bengal became Aryanised and adopted Hinduism. By the time Hiuen Ts'ang (Yuan Chwang) the famous Chinese Buddhist pilgrim, travelled through the area as late as the seventh century, more than half a millennium after the ruling house is thought to have claimed descent from Vishnu, 'Ts'ang' stated that 'the Aryanisation of the language does not appear to have progressed much.' (Clark 1969:191–192, citing Chatterji 1963:34).

The end of the 12th century saw the devastating arrival of the first Muslim invader Bakhtiar Khilji, who overran both the kingdoms of Pundra-Vardhana (with its capital at Gaur) and Kamrupa (with its capital in the vicinity of Guwahati). The Muslim invaders were soon repelled from Kamrupa, and despite further attempts during the 13th century, failed to gain any lasting control over the present-day regions of Rangpur, Cooch Behar and Jalpaiguri which all remained within Kamrupa society and polity. The area of Pundra-Vardhana, on the other hand, became from this time on permanently united with the

---

[4] Alternative views, exemplified by Ram Prasad Majumder (1955), are not generally shared by scholarship (cf. Clark 1969:175ff.).

regions to the south within a social and political entity that came to be Bengal. With repeated assaults upon Kamrupa from Gaur in the south-west, the capital was moved from Guwahati (in today's Assam) to Kamatapur (in today's Cooch Behar district), in order it would seem to better counter the aggressive Muslim neighbours. Acharyya (1966:150) writes that this shift in capital 'took place immediately after the defeat and death of Malik Yuzbeg [the third Muslim invader] in 1255'.[5]

Sociohistorical records are not explicit regarding the timing of the language shift from Tibeto-Burman to Indo-Aryan language; nevertheless we may surmise that a major factor in this regard was the shift to Kamatapur of the Kamrupa seat of government.[6] Prior to this, the capital had been located near Guwahati, and had been a driving force for Aryanisation of speech in the Brahmaputra valley (Clark 1969:197). The capital was the centre, both religiously and politically, from which Aryan influence radiated outwards. The shift of capital westwards to Kamatapur in the 13th century established a new centre of cultural influence, and must have given a great impetus to the Aryanisation of the Tibeto-Burman peoples living around Kamatapur. Geographically located at the heart of what would become the KRDS-speaking area, Kamatapur as capital would have been a point of social reference for the surrounding villages, and a force for social integration, to a degree that Guwahati as capital had not been, on account of its considerable distance to the east. I hypothesise that Indo-Aryan influence greatly increased in the KRDS area after the shift in capital and that this increased Aryanism gave rise to the proto Kamta SC and its language (defined by morphological changes 31, 32, 45, 63, 69 and 74). This proto Kamta SC and its language is the phylogenetic ancestor of all present-day KRDS lects.

Given the shift of capital as a *terminus post quem* for the linguistic Aryanisation of the peasant class in the KRDS-speaking area, the emergence of proto Kamta—distinct from proto Bangla and proto Asamiya—may be dated as *subsequent to AD 1255*. This is the 'leftwards bookend' after which the emergence of proto Kamta makes sociohistorical sense. The 'rightwards bookend', or *terminus ante quem*, is provided by *kingdom expansion* in the 16th century.

7.3.1.2 *Terminus ante quem*: kingdom expansion

At the present time, the lects defined by the six KRDS-wide changes are distributed as far west as Morang district of Nepal. This area goes considerably beyond the boundaries of the Kamata-Kamrupa kingdom of the 13th to 15th centuries, but corresponds neatly with the westernmost limit attained after the expansion of the Koch kingdom during the 16th century. The history is recorded as follows.

After the shift in capital to Kamatapur, the kingdom seems to have been referred to as *Kāmatā* rather than Kamrupa, and during the 14th and 15th centuries the rajas of Kamta were referred to as *Kāmatesvara* 'Lord of Kamta'. Achharya (1966:144ff.) lists three dynasties of kings who ruled Kamta during this period. The third of these dynasties was the Khyans or Khens, whose rule came to an end in AD 1498 with the onslaught wrought by Alauddin Hussain, the Sultan of Gaur (1493–1519). Kamatapur was sacked, and the

---

5 Accharya states elsewhere that 'We have no record of serious trouble from the Ahoms who were at this time establishing themselves firmly on the eastern part of Assam. The change of capital, therefore, probably had no connection with the Ahom invasion of Upper Assam' (ibid.: 143–144).

6 Nowadays, this area is completely Aryanised, though with pockets of Tibeto-Burman speakers still inhabiting the thick jungle areas of Jalpaiguri to the north of Cooch Behar.

Kamatesvara overthrown. Once again, the Muslim control over Kamta was short lived, and the occupying forces were driven away around AD 1505 by local chieftains called 'Bhuyans' with the aid of the Ahom king (Achharya 1966:177). A power vacuum ensued in the area, until a member of the Koch tribe known as Bisu (later given the more illustrious name of Biswa Singha) subdued the local chieftains one by one, and made himself 'the master of a dominion extending as far as the Karatoya in the west and the Barnadi in the east. He made a magnificent city in Kochbehar as his capital' (ibid.:189–190). The extent of this kingdom is shown in Figure 7.1. During the reign of Biswa Singha his followers began to be called *Rājbanshi* 'royal race or lineage' (Gait 1905:45), a term which grew in popularity in later centuries and pertains to this day.

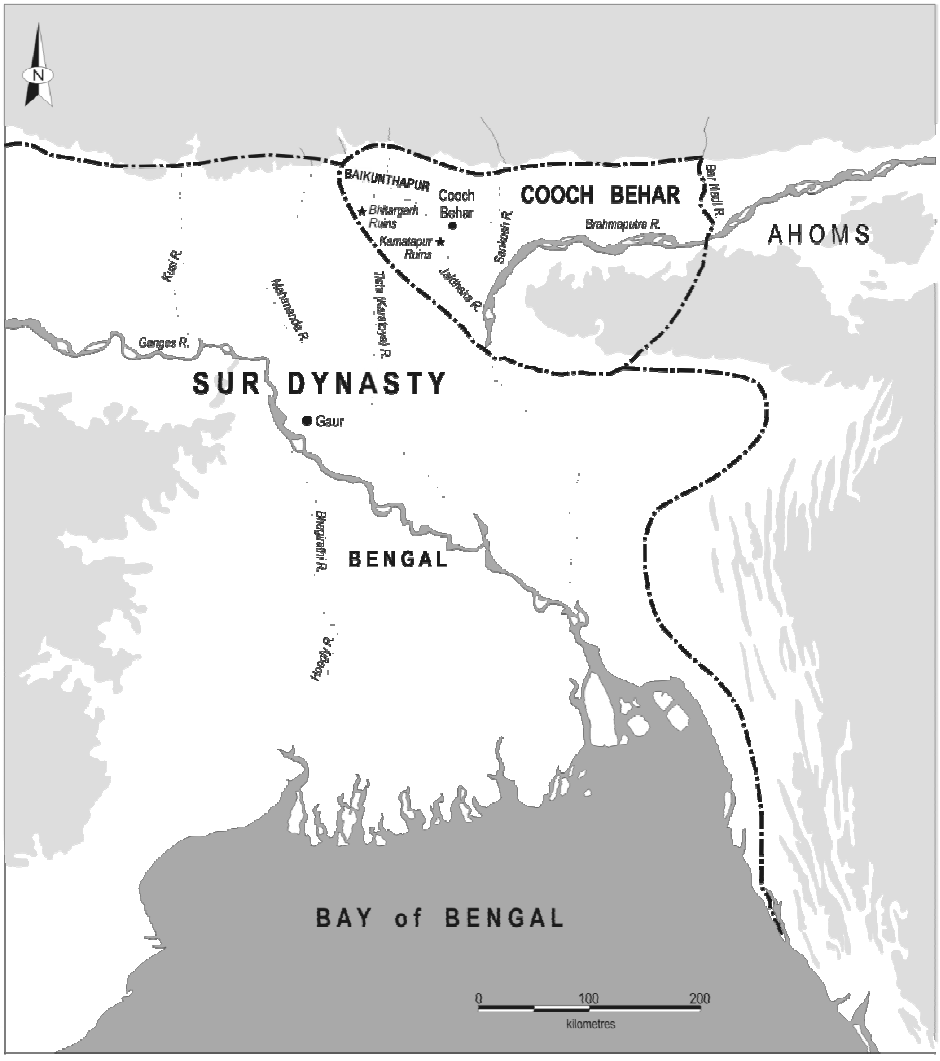


Figure 7.1: Cooch Behar under its first Maharaja, Biswa Singh c. 1540 (reproduced from Whyte 2002:546)

After Biswa Singha's death, his son Malla Deva became Maharaja of Cooch Behar and assumed the title Nara Narayan. Another of Biswa Singha's sons, Sukhladhvaj, became the commander-in-chief of the Koch armies (Gait 1905). Sukhladhvaj was an enormously successful general. He extended the boundaries of his brother's kingdom in all directions (excepting south-westwards to Gaur) by defeating the Ahoms and the Kacharis, the

Jayantia, Tippera and Sylhet kings, and winning the submission of other rulers (Acharyya 1966:194ff.). On account of his successes on the battle field the Maharaja's brother was nicknamed Chilarai 'the kite king'. The geographical extent of the Koch kingdom under Maharaja Nara Narayana and commander-in-chief Chilarai is shown in Figure 7.2.

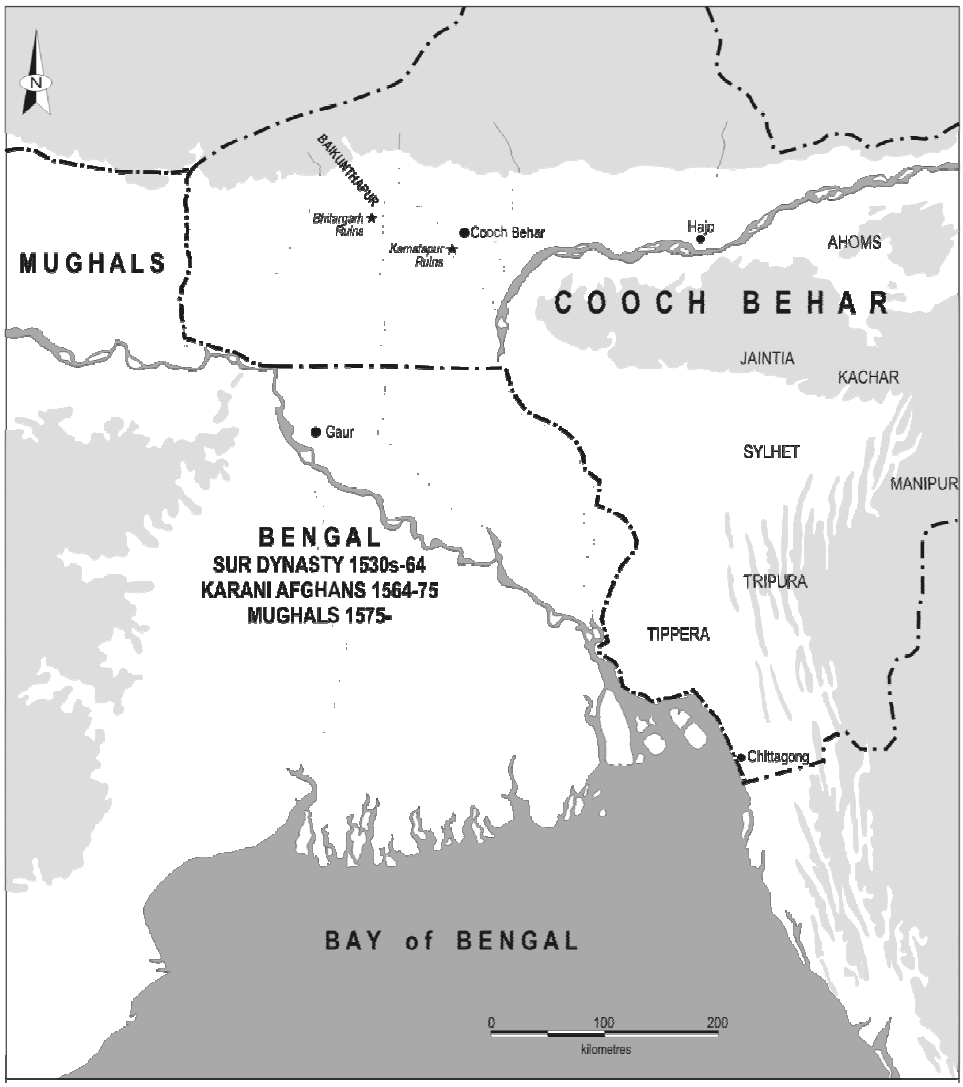


Figure 7.2: Cooch Behar's zenith under Nar Narayan (after Ahmed 1936) c. 1560 (reproduced from Whyte 2002:467)

These territorial gains, though substantial, were not long enjoyed. The Maharaja Nara Narayana lacked his brother's inclination or genius for war, and the death of Chilarai from small pox during a campaign against Gaur (c. AD 1571, Acharyya 1966:206) was followed by the gradual break up of the Koch kingdom.

From this account of 16th century sociopolitical history we can distil the following points relevant to sequencing KRDS linguistic history:

1. The 16th century expansion of the Koch kingdom is the most plausible sociohistorical explanation for the linguistic connection between today's Morang and Jhapa (in Nepal) and the rest of the KRDS area (in today's India and Bangladesh). It is plausible that this sociopolitical expansion resulted in the westwards migration of speakers of the old Kamta language (defined by its innovative features) into the newly conquered jungle area of today's Nepal Terai. As well as introducing new inhabitants to the area, the occupation of these territories would also have exposed its Tibeto-Burman-speaking occupants to the proto Kamta language. *Never before, and never since, did the control and influence of Kamta stretch so far west.*
2. However, the sociohistorical connection between Morang and Koch Behar was short-lived, and therefore it is unlikely that the proto Kamta innovations were propagated across this area after the expansion. Instead, it is more sociohistorically plausible that the proto Kamta morphological innovations underwent propagation *in the regions near Koch Behar after its Aryanisation, and before the expansion westwards*. The innovative features would then have been carried to the newly conquered western area as part of the proto Kamta language spoken by the Koch armies and settlers.

This sociohistorical scenario of *propagation followed by migration* accounts for:

(a) The presence of the diagnostic proto Kamta changes in the MH, RL and KS lects included in this reconstruction.

(b) The influence of Hindi and Bihari lects upon these same three lects; from the 17th century these areas fell beyond the bounds of Koch political control and social influence, and instead came under Mughal control and the social influence of Mithila in northern Bihar. It is notable in this respect that the great roads built by Maharaja Nara Narayana stretched *eastwards* from Koch Behar rather than *westwards* (Gait 1905:48, 56).

(c) The division of the proto Kamta SC into smaller SCs defined by more localised PEs. As stated above, the westwards expansion of Koch control and influence was not a lasting phenomenon. This increase in geographical separation between speakers, and absence of ongoing sociopolitical connection, would have resulted in *reduced interaction* between speakers in the west and speakers in Cooch Behar proper. Therefore, the most plausible sociohistorical effect which the political expansion had upon the proto Kamta speech community is the *division of its former unity*.

In summary: based on sociolinguistic reconstruction, the proto Kamta innovations seem to have been propagated within a PN sometime between the 13th and 16th centuries. The *terminus post quem* for this proto Kamta network was the shift in capital to Kamatapur during the 13th century, and the *terminus ante quem* was the expansion of the Koch kingdom during the 16th century. The proto Kamta speech community was a population of speakers (over several generations) who lived east of the Karatoya-Tista river in the region near Kamatapur, amongst whom specific and unique innovations ([MI 31], [32], [45], [63], [69] and [74]) were propagated during the early centuries of their Aryanisation. These common innovations have been inherited by the linguistic descendants of proto Kamta, which include the present-day KRDS lects and also at least some of the Hajong lects

spoken in the Garo hills. To date, no dialectological study has been made of the Hajong lects on which a reconstruction of historical connections with KRDS could be based.

7.3.1.3 Comparison with the Maharaja's letter of AD 1555

This overall hypothesis receives some confirmation from the letter of Maharaja Nara Narayana to the Ahom king, written in AD 1555. The following points of comparison may be made between the language of the letter, and the innovative proto Kamta features dated above as AD 1250–1550:

- [MI 69.] The agreement ending *-ɔn(t̪i) '2PL' is not found in the letter. The second person agreement endings used to address the Ahom king are *-ek* '2SG' and *-ā* '2SG'.[7] These do not prove the absence of *-ɔn(t̪i) '2PL' in the language of the time, but indicate that *-ɔn(t̪i) did not function as a marker of 'high honour second person singular' agreement.
- [MI 74.] The agreement ending *-ek* is found with *future* imperative function for second person singular in the letter. The overlap in function between general imperative (as is reconstructed for *-ɛkɔ '2SG.IMP' in §6.4.2.2) and future imperative (as is attested by *-ek* '2SG.FUT.IMP' in the letter) provides support for the reconstruction of [MI 74.] as a pre-16th century change. There are no instances of a present imperative formation in the letter.
- MI 63.] -u˜ '1SG', -ɔ˜ '1PL' in AGR.IIA. A pre-16th century sequencing of this change is also supported by the Koch king's letter. He writes **লেখনং** কার্য্যঞ্চ। এথা আমার কুশল। তোমার কুশল নিরন্তরে বাঞ্ছা **করি**। The crucial words here are given in bold and are, respectively, transliterated as *lekhnaṁm* 'we have written' (probably pronounced [lekʰnɔŋ]) and *kari* 'we do' (probably pronounced [kori]). The second form is included as evidence that when the letter writer is the grammatical subject, the agreement is with *first person plural*. The form [lekʰnɔŋ] conforms with the reconstructed secondary ending *-ɔ˜ '1PL' (here probably pronounced [-ɔŋ]).
- [MI 31.] The dating of this unique proto Kamta change prior to the 16th century is supported by the use of ইমৰাক *imrāk* 'them' in the Maharaja's letter.
- [MI 32.] The use of *imrā-* as an oblique pronoun (suffixed by *-k* 'DAT') in the Maharaja's letter raises some doubts as to whether [MI 32.] was a regular or variable part of the language in AD 1555. This reduces the diagnostic value of [MI 32.].
- [MI 45.] *-ʈʰɛ + kuna > *-ʈʰɛkuna 'place' as a base of locational pronominals. This change is not attested in the letter, but the letter is very short and the omission of *-ʈʰɛkuna 'place', as in the case of the 2PL ending *-ɔn(t̪i) discussed above, does not prove its absence from the language of the time.

---

[7] The presence of *-ā* is telling, and suggests the proto Kamta system reconstructed in §6.4 should be expanded to include *-a. This would also have implications for the reconstruction of proto Kamrupa, as the emergence of *-a '2SG.H' in secondary systems is plausibly a common KRDS and Asamiya (i.e. proto Kamrupa) change.

The full text of the letter may be read in Barma (1991:3–4). The above analysis of the language of the letter is not intended to be exhaustive; but this much is sufficient to verify that the Maharaja's letter supports the pre-16th century chronology assigned to [MI 74.], [MI 63.] and [MI 31.] by sociohistorical sequencing. The sequencing of just one change—[MI 32.]—may need to be reassessed in the light of the letter's evidence.

7.3.2 The proto Eastern Magadhan speech community

A complete reconstruction of the sociohistorical formation and breakup of the eastern Magadhan speech community—from which we get Bangla, Oriya, Asamiya, KRDS, plus others—is beyond the scope of this study. Nevertheless, a few points are in order to clarify some misconceptions that have been held in this regard.

Firstly, consider the innovative features given by Chatterji as characteristic of the eastern Magadhan lects (1926:93–94):

- Fully rounded pronunciation [ɔ] of the character written as অ, অ, অ (transliterated as '*a*'). This contrasts with the inherited unrounded pronunciation of the NIA midlands.
- A palatal pronunciation of the inherited sibilant. However, because this pronunciation is inherited from proto Magadhan, and not innovative, it is not diagnostic of a propagation event that would define eastern Magadhan.
- Transposition of medial *i, *u is a regular feature of these lects, unlike in the Bihari lects where it is a sporadic occurrence.
- *kṣ* (in Tatsama words) becomes *(k)khy. However, this change is not unique to these lects (cf. Masica 1991:201).
- Genitive in *-r* < *-kēra, kara*. While this may have been a morphologically specific change (and hence diagnostic of eastern Magadhan) it is not of great diagnostic value given the general MIA pattern of leniting and deleting single intervocalic stops (cf. Masica 1991:180ff.).
- Disuse or restricted use of the genitive in *-ka*. The principle adopted in this study is that instances of disuse and loss of parts of constructions are of low value for diagnosing a PE (see further §3.4.1.1).
- Past and future bases in *-il, -ib*. This contrasts with the rest of the Magadhan lects which have *-al, -ab*. According to Chatterji's (1926:940ff.) reconstruction, the variation in vowel between eastern and other Magadhan lects is the result of inherited variation with subsequent regularisation. Chatterji reconstructs the etymology thus: 'the past base in «-il-» in Bengali, Assamese and Oṛiyā, in «-al-» in the Bihārī speeches, and in «-il-, -al » in Marāṭhī, and similar «-l-» forms in the other NIA. speeches, originated from the OIA. «-ta, -ita» *plus* the OIA. diminutive or adjectival affix «-la-» in the extended forms «-ila, -a-la, > -illa (-ĕlla),-alla»' (ibid.:941). Regularisation of inherited variation is not considered diagnostic of a PE in this study because of the possibility that the regularisation occurred independently and in parallel (cf. §3.4.1.4).
- A passive participle in *-ā*.
- 'confusion between roots « ah » and « hō » both meaning *is*, but derived from different roots'.

This evidence is a mixed bag: some is solid enough; some is easily dismissed. Certainly there is scope for fresh studies of the historical relations between eastern and other Magadhan lects, based on more stringent methodology of historical reconstruction.

Secondly, accepting for a moment that there is at least some evidence for a unified stage of development among the eastern Magadhan lects, we should consider the possible chronology of this unified stage. Chatterji is once more our clearest guide on the matter. Regarding the emergence of Oriya, he writes:

> In the early part of the 7th century, we have … the testimony of the Chinese traveller [Hiuen Ts'ang] that the sea-board country where Oṛiyā is now spoken was non-Aryan in speech. Yet we have epigraphical evidence to show that Brahmans were settled in non-Aryan Kōŋgōda with grants of land precisely when Hiuen Thsang noticed the general linguistic condition of the country … What would seem to have been the case is that the Ōḍra people were receiving Aryan speech from the neighbouring Suhman and Rāḍha, in the 7th century and before, as well as during the subsequent period, and they rapidly became Aryanised (Chatterji 1926:105–106)

The first point to be made from this statement is that a pre-Oriya lect emerged, distinct from pre-Bangla, in all likelihood *only after the 7th century*. This is the *terminus post quem* for the division of any proto eastern Magadhan speech community. There is evidence (again from Hiuen Ts'ang) that Aryanisation of language had begun (though not progressed very far) in Kamrupa by the 7th century—at least in the regions nearby the capital (in the vicinity of Guwahati).

Thirdly, we may consider the mechanisms for propagation of changes that may have existed during late MIA and very early NIA. During this period Indo-Aryan language and culture was spreading into the areas of Odra and Kamrupa and Banga, but the spread was patchy and the bulk of the general populace probably remained non-Indo-Aryan speaking. As mentioned in §7.3.1, migration and resettlement of Brahmans was not at all uncommon, and indeed became more frequent as more areas embraced Aryan religion. In all likelihood it was this migration and interaction between the priestly classes which enabled propagation of common eastern Magadhan linguistic innovations between the outlying colonies of Aryanism in Kamrupa, Banga and Odra, as they lived surrounded by non-Aryan speaking peoples.

In conclusion, the case for a unified eastern Magadhan stage, characterised by propagation events, remains to be made using robust historical methodolgoy. Some evidence has been put forward by Chatterji, but renewed research on this matter is required. The situation may turn out to be similar to how Grierson described it:

> East of Māgadha lay the Gauḍa or Prāchya Apabhraṁśa, the head-quarters of which were at Gaur, in the present district of Malda. It spread to the south and south-east, and here became the parent of modern Bengali. Besides spreading southwards, Gauḍa Apabhraṁśa also spread to the east keeping north of the Ganges, and is there represented at the present day by Northern Bengali and, in the valley of Assam, by Assamese. Northern Bengal and Assam did not get their language from Bengal proper, but directly from the west. Māgadha Apabhraṁśa, in fact, may be considered as spreading out eastwards and southwards in three directions. To the north-east it developed into Northern Bengali and Assamese, to the south into Oriya, and between the two into Bengali. Each of these three descendants is equally directly connected with the common immediate parent, and hence we find Northern Bengali agreeing in some respects rather with the Oriya spoken far away to the south than with the Bengali of Bengal proper, of which it is usually classed as a subordinate dialect. (Grierson 1903–1928:vol.1:126.)

Suffice it to say here that this general picture would need, at the least, to be supplemented with accounts of the later common innovations which were propagated across these partially differentiated lects (see below).

7.3.3 Considering the proto Bengali-Assamese speech community

The KRDS lects are classified in the Ethnologue, along with 15 other lects, as Bengali-Assamese (Gordon 2005). This has long been held to constitute a historical linguistic subgroup. 'The agreement between Assamese and Bengali is so close that the dialects of Bengali and Assamese may be described as belonging to the same group' writes Chatterji (1926:108)—he does not, however, give common innovative features as proof of this point.

Chatterji's subgrouping hypothesis has been subjected to detailed comparative reconstruction by Pattanayak (1966). In his historical phonological study, Pattanayak provides four innovations that define the putative proto Asamiya-Bangla stage:

(1) merger of alveolar and postalveolar nasals;
(2) merger of alveolar and postalveolar laterals;
(3) merger of *ɖ and *ɖ^{ɦ} in medial position;
(4) lowering of *e > /ɛ/ when followed by a low vowel.

The first and second innovations have been discussed in §4.3.7, where they are found to have occurred subsequent to the 15th century and include not only Bangla, Asamiya and KRDS, but also Hindi and the Bihari lects among others. The third of Pattanayak's innovations is the merger of medial *ɖ and *ɖ^{ɦ}. This innovation has been discussed in §4.3.4. Out of the eight KRDS lects included in this reconstruction, it is regular in only SH, RP and BH, with some irregularity in BN. It is part of a broader phonological change affecting medial *h, *m^{ɦ}, *n^{ɦ} and *l^{ɦ}, turning the breathy voicing into modal voicing (see [PI 9.]). The chronology of this change is reconstructed tentatively as post-16th century. Chatterji writes:

> The aspirates, initial and intervocal, which Bengal inherited from OIA, were preserved intact in the [Old Bengali], and to a very large extent in the [early Middle Bengali] period. But even from the [early Middle Bengali] period, from the latter part of the 15th century it would seem, (judging from the orthography of Early Bengali MSS., and from [New Bengali] history of the aspirates), the aspirates as well as «-h-» grew rather feeble in an intervocal position—and also finally (Chatterji 1926:441).

And again, addressing specifically the *ɖ/*ɖ^{ɦ} merger:

> It seems in the early 16th century, voiced aspirated forms like পড় «påṛh-» *read* ... বাড়ে «bāṛh» *increases* … still obtained, although it is likely that the aspiration had become feeble. The voiced aspirates seem to have preserved the aspiration (in the West Central dialect [i.e. SCB—MT]) longer than the unvoiced ones, in both final and intervocal positions. (Chatterji 1926:442)

The fourth of Pattanayak's changes pertinent to Bangla-Asamiya historical relations, is lowering of *e > /ɛ/ when followed by a low vowel. This seems to be a misinterpretation of the correspondence sets. The historical event was not the lowering of *e to /ɛ/ but, as argued in §4.4.3, the phoneme /ɛ/ is the inherited phoneme in these lects, with /e/ resulting from loan words and regressive raising of *ɛ to /e/ before a high vowel. In Bangla /e/ also results from the lowering of *i, but not in Asamiya. As articulated in §4.4.3, the presence

of regressive raising *ɛ > i in early Oriya suggests that regressive raising of *ɛ is inherited from the eastern Magadhan stage (though this stage is yet to be adequately reconstructed). Consequently, we cannot identify a unique Bangla-Asamiya stage of historical interaction based on either of the phonemes /e/ or /ɛ/.

In summary, none of Pattanayak's changes are diagnostic of a unique proto Bangla-Asamiya subgroup that also includes proto Kamta. Pattanayak's first three innovations are post-proto Kamta events which spread across already differentiated lects after the 15th century. The sociohistorical conditioning and subgrouping value of that diffusion across Bangla, Asamiya and part of KRDS is considered in §7.4.4. Grierson's contention may well be true that 'Gauḍa Apabhraṁśa' was the parent speech both of Kamrupa and today's Bengal (see quote under §7.3.2), *but it has not yet been proven as such by careful historical linguistic reconstruction*. Statements to the effect of 'but it's obvious' do not qualify as historical arguments as they do not properly differentiate between retentions and innovations, or, between *general NIA* propagation events which are post-15th century (see §7.4.4) and the pre-13th century speech communities that were already differentiated before the later common changes came into effect.

Though it has not been the purpose of this study to reconstruct higher level protolanguages beyond proto Kamta, the reconstruction here has turned up three morphological innovations—[MI 73.] (diagnostic), [MI 2] (supportive), [MI 70] (supportive)—which provide some evidence for a protolanguage which may be termed proto Gauḍa-Kamrupa. Furthermore, [MI 40.] may also have undergone propagation during the proto Gauḍa-Kamrupa stage, with subsequent replacement of *mɔn̪ɔ 'like, similar to' through change [MI 38.] in north-west KRDS. (North-west KRDS is defined in §7.5.2).

7.3.4 Considering the proto Kamrupa speech community

The *Kamta-Asamiya* subgrouping hypothesis was probably first articulated by Grierson (see quote at the end of §7.3.2). At this point, Chatterji (1926, cf. p.148) and Kakati (1962, cf. p.5) concur with Grierson's diagnosis, and the same position is reflected in recent statements, like that of Baruah and Masica (2001:43):

> Assamese is most closely related to Bangla, particularly to the northern dialects of that language. In fact, the (northern) Rajbangshi 'dialect' of Bangla could be considered a dialect of Assamese that has come under Bengali cultural hegemony.

This view entails (in traditional terms) a subgrouping of KRDS and Asamiya, and (in the terms of this study) a common ancestral Propagation Network diagnosed by a PE.

The problem with this proposition is that, as for the Bengali-Assamese subgroup, it is yet to be substantially proven. While stated in several previous academic works, none that I have found present *linguistic evidence* in its support. Instead, the position is mostly argued on the basis of political history, including the king of Cooch Behar's patronage of early Asamiya language, among other politico-historical links between Kamta and Kamrupa. The position taken by this study is that political history (and social history more generally) is relevant to linguistic history and phylogenetic classification *only in so much as it conditions the propagation of linguistic innovations among speakers*. In the absence of unique and diagnostic innovations, *political or social history is not relevant to linguistic history*.

The use of political history in this chapter is thus unlike the use of political history in, for example, the following quote from Kakati (1962:6):

> It was under the patronage of kings outside the western limit of modern Assam,—under the patronage of the kings of Kāmatāpur … that the earliest Assamese books were written. Even now the spoken language of North Bengal and western Assam (districts of Kāmrūp and Goālpārā) is substantially the same and seems to form one dialect group.

The only evidence for this one-ness of dialect grouping that Kakati presents is the early medieval political unity. Kakati's argument does not constitute a historical *linguistic* argument because it does not centre on the distribution of unique and diagnostic linguistic changes.

What evidence then is there of *linguistic innovations* common to all KRDS and Asamiya lects, but not common further afield in NIA? Are we justified in talking of Asamiya and KRDS as a subgroup? The phonological and morphological reconstruction of the present study has found three morphological innovations that give some answers to these questions:[8] [MI 67.] (diagnostic), [MI 22.] (supportive), and [MI 23.] (supportive). These changes provide evidence for a proto Kamrupa stage of linguistic history—ancestral to proto Kamta and proto eastern Kamrupa (Asamiya). However, a thorough KRDS-and-Asamiya-wide reconstruction of linguistic history is required before this protostage can be robustly established.

The question remains: did the propagation of [MI 67.], [MI 22.] and [MI 23.] occur before or after the propagation of the changes that define the proto Kamta stage? A plausible, yet perhaps inconclusive chronology is suggested by the sociohistorical account given in §7.3.1 for the origin of the proto Kamta speech community.

> If Indo-Aryan language was established first in the area surrounding the Kamrupa capital, and only became established in today's KRDS area after the shift of capital to Kamatapur;
>
> Then it is plausible, though not conclusive, that [MI 67.], [MI 22.] and [MI 23.] had occurred during the period of the eastern capital of Kamrupa (pre-13th century), and that these changes were inherited into the proto Kamta language after the shift in capital.
>
> This hypothesis suggests a common linguistic origin for both KRDS and Asamiya in an earlier proto Kamrupa speech community.

There are further *distinctive* phonological and morphological features shared between KRDS and Asamiya, but such features tend to be either *inherited* from Eastern Magadhan, or instead, propagated subsequent to the breakup of proto Kamta and thus *only reflected in a subsection of KRDS*. The distinctive but non-innovative features common to KRDS and Asamiya include the case markers: *-ɔkɔ 'dative-accusative' and *-ɔṭ̪ɔ 'Locative'. See §5.3.6 and §5.3.7 for the arguments that these case markers are inherited from a pre-Kamrupa source and as such are not diagnostic of proto Kamrupa.

The history of innovative features shared between Asamiya and only a subset of KRDS lects is analysed further in §7.5.4.2.

---

[8] There is plenty of evidence for major linguistic propagation between Asamiya and *eastern* KRDS (Bongaigaon), see §7.5.4.2.

7.3.5 The proto Asamiya speech community: 13th century onwards

There are many innovative phonological and morphological characteristics of the Asamiya lects. Kakati (1962) is the primary exponent on this topic, with his main intention to establish that Asamiya and Bangla are structurally so removed that they should not be considered dialects of the same language. This is an essentialist definition of 'a language', such as has been critiqued for historical purposes in Chapter 3. Unsurprisingly, the distinctive features he outlines for Asamiya in contrast with Bangla fail to distinguish between inherited and innovative features. Therefore, not all of the features he ascribes to Asamiya are diagnostic of the proto Asamiya stage of linguistic history, but instead merely diagnose its Magadhan inheritance. The phonological and morphological features of Asamiya outlined by Kakati (1962:8ff.) are as follows:[9]

- Asamiya 'follows the pan-Indian system of penultimate stress and Bengali has initial stress'. The Asamiya feature is not an innovation.
- The genitive case affix is /-er/ in Bangla and /-ɒr/ in Asamiya, with /-er/ maintained in an instrumental case suffix in Asamiya. This is a case of inherited variation with subsequent regularisation (cf. §3.4.1.4) and hence is not diagnostic of a propagation event (cf. §5.3.5).
- The locative affix in Asamiya is /[ɒ]-t/ in contrast with Bangla /-ṭe/. The innovation involved here is on the part of Bangla not Asamiya (see §5.3.7) and so this change is relevant to defining proto Bangla, not proto Asamiya (cf. §7.3.6 below).
- The present participle in Asamiya is /oˇṭ/, Bangla has /-iṭ-/. Early Asamiya has *-ante*. 'The Pres. participle in Oṛiyā is *-anta-*, and both the [Asamiya] and [Oriya] forms go back to O.I.A. and M.I.A. active participle in *-ant-*' (Kakati 1962:361). This feature is inherited, and not diagnostic of an Asamiya propagation event. Incidentally, Chatterji reconstructs the Bangla suffix as derived from the same MIA source (1926:999), with the /i/ element the result of transposition of an earlier feminine suffix (Chatterji 1926:653–654).
- A past conditional conjugation in Asamiya is 'expressed by the postposition ɦɛˇtɛn, *heˇten* (earlier *haˇte, hante*) after a fully conjugated verbal root in the past. Bengali expresses the past conditional with the pres. part. base in *-it* with personal conjugational affixes' (Kakati 1962:9). That is, the difference is between the construction of conditional sentences in Asamiya and Bangla. In Asamiya the consequent clause, or apodosis (and optionally the conditional clause, or protasis) is followed by /hɛˇṭɛn/. For example, in modern Asamiya (from Kakati 1962:360): /tumi kɔle xi ahil hɛˇṭɛn/ 'had you said, he would have come'. An example from early Asamiya writings (from ibid.):

> jadi āji gharata āchila hante svāmī, tebe āni tomāka rākhilo hante āmi; (Daityāri: Śankara Carita) 'If my husband had been at home today, I would have taken you in and kept you'.

---

[9] Kakati also lists some lexical distinctives of Asamiya in contrast with Bangla. However, this study does not include reconstruction of lexical innovations, and therefore it is beyond the scope of the present reconstruction to evaluate whether these lexical features constitute innovations, or features retained from an earlier Magadhan stage.

Indeed this conditional construction using *hante* > /hɛ̃t̪ɛn/ is innovative, and distinct from functionally similar constructions in both KRDS and Bangla. Therefore *this* distinctive feature is diagnostic of an earlier stage of Asamiya linguistic history.

- The infinitive affix in Asamiya is derived from /-ib/, while in Bangla it is derived from /-it̪/. The Asamiya morpheme is inherited in this function—cognate with KRDS and Oriya (cf. §6.2.2)—and not diagnostic of proto Asamiya.
- Asamiya 'has a complete set of negative conjugation with the negative particle nɒ, *na-* placed before the verb root. Oriya has a negative conjugation with the verb substantive only. Bengali has no negative conjugation' (Kakati 1962:10). Going on cross-referenced portions of Kakati's work (ibid.:383), what seems to be in view here is that in Asamiya the negative morpheme is always found before the verb, in contrast with Bangla and Oriya (and most of KRDS) where this morpheme generally follows the verb. This indeed could have been a propagated innovation, but the contact-related motivation for this change reduces its value for diagnosing a single propagation event (cf. P.C. Bhattacharya 1975).
- The plural nominative suffixes in SCA are distinctive and unique. However, they are not shared with western (Kamrupi) Asamiya (U. Goswami 1970), and thus are not diagnostic of a common proto Asamiya stage of linguistic history. The western Asamiya plural suffixes are instead cognate with KRDS and Bangla forms (cf. §5.4.3).
- Asamiya 'pronominal derivatives of time and place seem to have no parallel formations in Bengali'. At the very least, the pronominal derivatives of place discussed by Kakati (ibid.:325ff.) do seem to constitute an innovative use of the locative affix within this pronominal set.
- In Asamiya, *a > ɒ / _Ca. This process of 'shortening' of nonstressed *a is very similar to that described in §4.4.7 for western KRDS lects, and also found in Hindi and Bihari. With these connections further afield, it is unlikely that this change is diagnostic of a proto Asamiya propagation event.
- The distinctive symbol ৱ for the /w/ glide found in Asamiya has no counterpart in Bangla. This is an orthographic, and not a linguistic feature. It has no bearing on historical linguistic relations.

As editor to the second edition of Kakati's work, Golok Chandra Goswami outlines further distinctive features of Asamiya in a footnote to Kakati (1962:8), which are here paraphrased in the terms of this study:

- The alveolarisation of dental and postalveolar stops, resulting in the merger of these two series. This is a diagnostic innovation (cf. §4.3.6).
- The spirantisation of the laminal affricates (i.e. palatal series). The diagnostic value of this change is doubtful (cf. §4.3.9).

- Changes to the inherited sibilant (see §4.3.13). These changes are common with the eastern Bangla varieties, and seem to be related to contact with Tibeto-Burman varieties. As ecologically nondistinctive changes, they are not diagnostic of PEs (cf. §3.4.1.2).

Based on this critical evaluation of the evidence put forward by Kakati and Goswami in Kakati (1962), there are at least four changes (three morphological and one phonological) which diagnose a common Asamiya stage of historical development:

(1) the innovative past conditional construction
(2) the prefixed negative particle
(3) the locational pronominals
(4) the merger of dental and postalveolar stops

None of these changes are common with KRDS beyond Bongaigaon. It seems therefore that they were innovated after the split of proto Kamta from proto Kamrupa in the 13th century. For the purpose of this study it is not necessary to be any more specific regarding the chronology of the proto Asamiya changes, other than to establish that they are phylogenetically parallel to the proto Kamta changes described in §7.3.1.

7.3.6 The proto Gauḍa-Baŋga (Bangla) speech community

Similarly to the grouping of Asamiya with KRDS, the Bangla lects have been grouped together for primarily historical-political rather than historical-linguistic reasons, with little evidence presented by way of common linguistic innovations. The key historical reconstruction is Chatterji (1926:146–147) who writes:

> Political and social reasons have brought about the present unity of speech in Bengal, despite the fact of dialects. From the time of the Pālas [of Gauḍ], the greater part of Bengal formed portions of one empire. Gauḍa and Vaŋga are frequently spoken of together, Gauḍa meaning North Central Bengal [south and south-west of Kamata and Kamrupa—MT], and the Western part of the Delta, and Vaŋga including not only Bengal beyond the Brahmaputra, but also a considerable part of the Delta. … If it had not been brought about by some sort of political union under the Pālas just when the foundations of the Bengali language were laid, and by the dispersion of a well-organised Brahman community all over Bengal, and Kāyastha participation in their efforts, the evolution of a common nationality and of one type of culture and literature among the people of heterogeneous origin in West Bengal, in East Bengal, in North Bengal, would have been extremely problematic

Chatterji postulated there to be four dialects ‘of Bengali’: Rāḍha, Varêndra, Vaŋga, and Kāmrupa. His tabulation of the historical relations between these lects (ibid.:140) is reproduced in Figure 7.3.

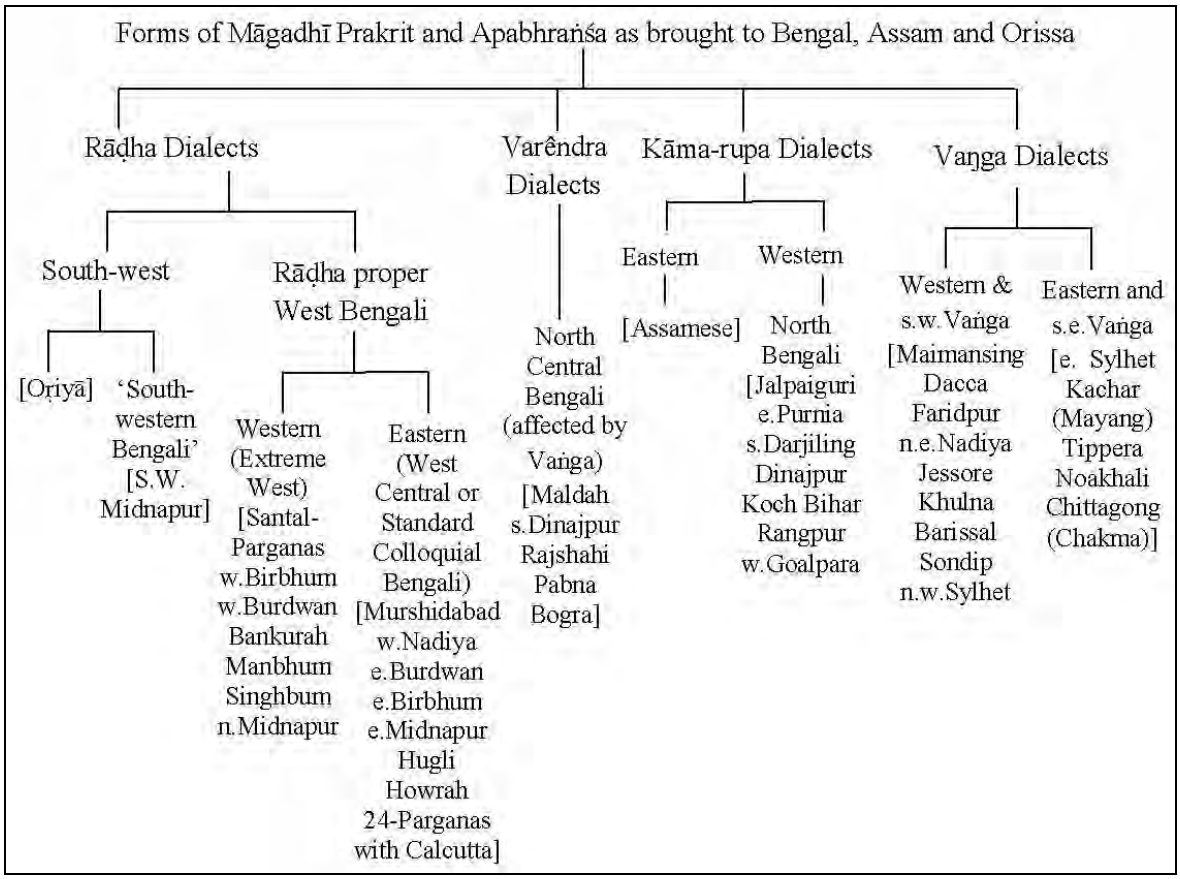


Figure 7.3: Chatterji's (1926) tabulation of the historical linguistic relations between eastern Magadhan lects

Note that Chatterji's classification of 'Bengali dialects' includes lects ancestral to both Asamiya and Oriya. However, Chatterji does not intend to classify these lects as dialects of Bangla. Therefore, Chatterji's four dialects—Rāḍha, Varêndra, Vaŋga, and Kāmrupa—should not be termed 'dialects of Bengali' but rather, 'dialects [in the sense of historical derivatives] of eastern Magadhan' (cf. Chatterji 1926:92ff.).

The question that remains to be answered is this: did Chatterji articulate a set of innovations that are diagnostic of a proto Bangla speech community and its language? (Or, to be more historically accurate, a proto Gauḍa-Baŋga SC, see below). The position Chatterji articulates on the origins of Bengali is summarised in the following paragraph:

> The Bengali dialects cannot be referred to a single Primitive Bengali Speech, but they are derived from various local forms of late Māgadhī Apabhraṅśa, which developed some common characteristics that may be called pan-Bengali: e.g. «-ila, -iba » for the past and future base, rather than «-ala, -aba » ; «-iā » rather than simple «-i » for the conjunctive; «-ēra < -kēra » besides «-ara < -kara » for the genitive ; «-kē, -rē » for the dative, rather than «-ku » as in Oṛiyā: etc. These pan-Bengali features link the dialects together as members of a single group, and enabled them to be attached to a composite literary language as a matter of course (ibid.:139).

Here Chatterji gives a set of features, which if they did 'link the dialects together as members of a single group' would constitute evidence for a proto Gauḍa-Baŋga stage of linguistic history. The putative 'pan-Bengali' features are examined in turn:

- /-il/ ‘past’ and /-ib/ ‘future’. These characterise eastern Magadhan as a whole, not just Bangla, and hence are pertinent to §7.3.2 and proto eastern Magadhan, rather than the reconstruction of proto Bangla per se.
- /-ia/ ‘conjunctive’, or ‘perfect participle’. This is found also in KRDS, early Asamiya and possibly early Maithili (cf. §6.2.1). Analogously to the situation in early Asamiya, Jha (1985 [1958]:513) gives *iā* as one of several variant forms found as conjunctive in early Maithili literature, though it has been replaced in the modern language. It seems doubtful therefore that the extension of inherited *-i by *-a is uniquely diagnostic of a proto Bangla (Gauḍa-Baŋga) stage.[10]
- /-er/ ‘genitive’. As argued in §5.3.5, the presence of this form is a case of *inherited variation with subsequent regularisation.* Such changes are not diagnostic of a propagation event (cf. §3.4.1.4).
- /-ke, -re/ ‘dative’. This extension of the inherited dative suffix by the locative-instrumental has been discussed in §5.3.6, and is supportive, rather than diagnostic of a propagation event.

None of these features can do the job for us of diagnosing a proto Gauḍa-Baŋga speech community and language, because none of the innovations pass the diagnostics developed in this study to sift integrated propagation events from cases of parallel independent development. However, certain diagnostic innovations unique to Bangla have shown up in the course of this study of KRDS, and no doubt several more would emerge from an in-depth comparison of the Bangla lects. These changes are not shared with KRDS. The diagnostic changes identified by this study are:

[MI 1.] > /-ḍ̲-/ ‘PL.OBL.AN’ {SCB} (before AD 1500). Diagnostic.

[MI 3.] /-[e]ra/ ‘PL.NOM’ in pronouns > /-[e]ra/ ‘PL.NOM.AN’ in general nominal morphology {SCB} (by the 15th century). Diagnostic.

[MI 7.] *-[ɔ]ṭ̲ɔ ‘LOC’ + *-ɛ ‘LOC-INS’ > /-ṭ̲e/ ‘LOC’ {SCB, Lodha} (before AD 1400). Probably diagnostic.

The textual evidence given by Chatterji points to a dating of these innovations before AD 1500 in the case of two changes, and before AD 1400 in the case of the other (cf. Chapter 5). As such, these innovations are diagnostic of a propagation event which is phylogenetically parallel with the proto Kamta and proto eastern Kamrupa (Asamiya) stages of linguistic history (reconstructed in §7.3.1 and §7.3.5 respectively). We lack as yet a *terminus post quem* for the proto Gauḍa-Vaŋga stage of linguistic development.

The protospeech community and language diagnosed by these innovations is here termed as ‘proto Gauḍa-Baŋga’ based on Chatterji’s observation that, historically, ‘Gauḍa and Vaŋga are frequently spoken of together’ (1926:146; Vaŋga can alternatively be Romanised as Baŋga which better shows the connection with modern ‘Banga-la’). The term ‘Bangla’ and ‘Bengali’ are more recent terms to denote the language and speech of

[10] This feature is a good illustration of Croft’s argument for why an essentialist, or structural-based definition of ‘a language’ will not do for linguistic phylogeny (cf. §3.2.2). The SCB (standard colloquial Bangla) itself no longer has its conjunctive in /-ia/, having been fused into /-e/. If ‘Bangla’ were defined (on an essentialist model) in terms of this feature, even SCB would no longer qualify as ‘Bangla’!

the region. Indeed up until the 19th century it was more common for this language and its dialects to be referred to as 'Gauḍa' than 'Bangla' (Chatterji 1926:148–149).[11]

7.3.7 The division of proto Magadhan

It has been argued above that several stages of the linguistic history of eastern NIA—which previously have been taken for granted—are yet to be thoroughly established by sufficiently broad and methodologically robust historical reconstruction. Further historical studies need to take up the task of reconstructing propagation events and speech community events for proto Kamrupa (the putative ancestor of Asamiya and KRDS), proto eastern Magadhan (the putative ancestor of Bangla, Oriya, Asamiya and KRDS as well as others), and indeed proto Magadhan itself. This study has taken the broadest of these subgroups for granted—proto Magadhan—and has reconstructed the divergence of proto speech communities and their languages out of this parent speech. The results are shown in Figure 7.4 in accordance with the adjusted phylogenetic tree model outlined in §3.4.4. In this figure, 'p' as in 'pKamta' means 'the proto speech community and its language' as defined in §3.2 in terms of PEs that resulted from speaker interaction. '?pKamrupa'—which is shown as the ancestor of both proto Kamta and proto eastern Kamrupa (Asamiya)—is prefixed with a question mark—as is '?pGauda-Kamrupa'—to remind the reader that these hypothetical stages are less robustly established in this reconstruction than the other proto speech communities.

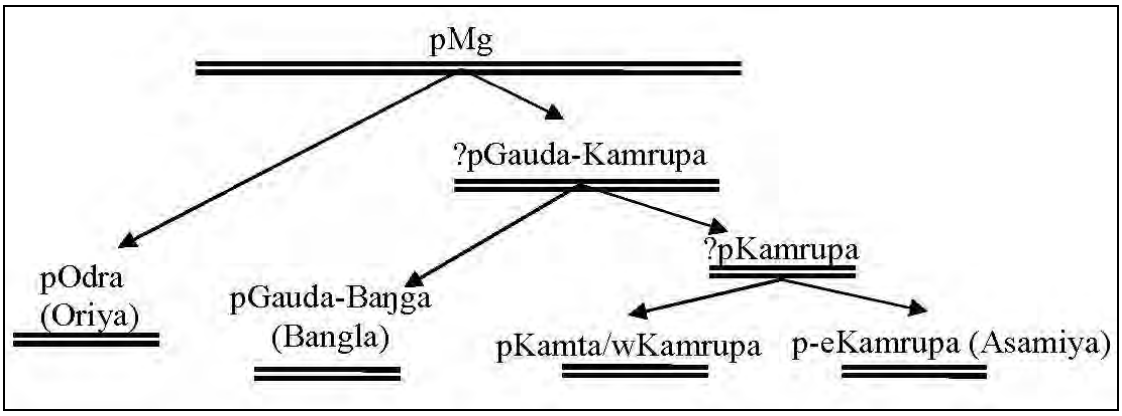


Figure 7.4: The historical fragmentation of proto Magadhan into its linguistic descendants

Notice that proto Gauda-Baŋga, proto Odra (Oriya) and proto Kamta did not undergo major phonological innovations during this early NIA period, though proro eastern Kamrupa (Asamiya) did. The protostages and phylogenetic relations for this period of linguistic history are diagnosed on the basis of morphological or morphologically-conditioned changes. The historical scenario reconstructed above (so far just up to AD 1550) is of the fragmentation of the Magadhan speech community and its language into several proto speech communities and protolanguages, with proto Kamta one of the historical linguistic fragments.

---

[11] Note that *Baŋga* is a romanised transliteration of বংগ and the vowels are pronounced [ɔ] not [a], thus [bɔŋgɔ].

7.4 The middle KRDS period: approximately AD 1550–1787

From here our story focuses on the history of KRDS lects, and the other Magadhan lects enter the discussion only as they contribute to PEs in KRDS linguistic history.

Proto Kamta took its inheritance from ?pKamrupa (and before that from ?pGauda-Kamrupa), innovated the unique features outlined above during AD 1250–1550, and then split into three main sections (western, central, eastern) as will be shown below. This division of the speech community was plausibly caused by the rapid sociopolitical expansion, and then fragmentation, of the Koch-Kamta kingdom in the 16th century (see above). The Koch-Kamta kingdom did not long maintain its supremacy over the conquered areas. The Mughal empire was expanding from the west through Bihar, and from the south through Bengal. To the east the Ahom kingdom had been re-established and control of Kamrup and Goalpara was to change hands several times between the Ahom and Koch kingdoms.[12]

The period AD 1550–1787 is marked by two concurrent phenomena: (i) local innovations which define distinct and localised propagation events, and (ii) wider changes, common to many NIA lects, which also spread across the KRDS area. These two sorts of PE—local versus wide-range—and the different SCEs that caused them, are reconstructed in §7.4.1–§7.4.4. In order to aid the reader in following the arguments below, the tree model of this period of KRDS linguistic history is given in Figure 7.5, in advance of the argument. The dotted line indicates the propagation of [PI 9.] across proto Gauda-Baŋga, proto cKRDS, etc., but not across proto wKRDS. Proto wKRDS is prefixed by a question mark to indicate that this stage is not well established at present. Arrows indicate the chronological relations reconstructed in this study.

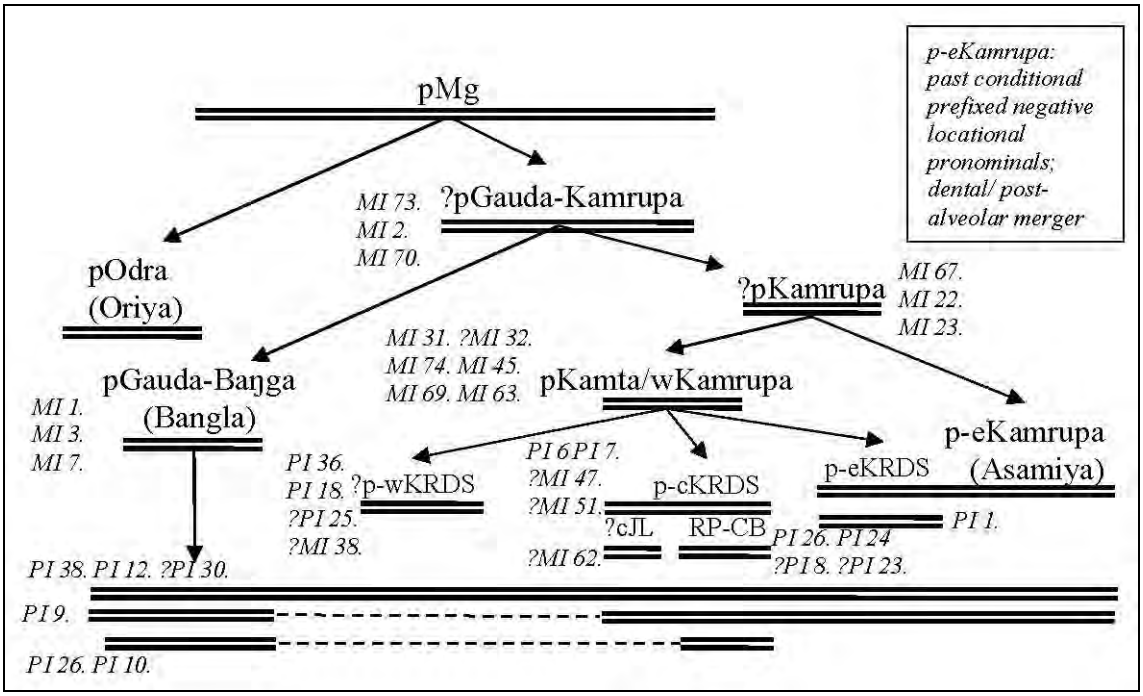


Figure 7.5: The division of proto Kamta, and re-integration with some other NIA lects (AD 1550–1787)

[12] Kamrupa [kamrupɔ] > Kamrup [kamrup] due to loss of final vowel, cf. [PI 38].

The intended meaning of this diagram, to be defended below, is that the proto Kamta SC and its language was split into three distinct SCs which underwent more localised PEs. The easternmost SC (here labelled p-eKRDS) underwent innovations in common with p-eKamrupa (Asamiya) to the east, and by this propagation event established an ongoing pattern of closer phylogenetic relations with Asamiya than with the other KRDS lects. The other two descendants of proto Kamta underwent unique local innovations, as well as participating in changes that had a wider range and significance in NIA. For example, proto cKRDS underwent [PI 9.]—the change of breathy to modal voicing in sonarants—in common with Bangla, Asamiya and Oriya. During this same period, all of KRDS underwent the loss of final *ɔ, which is a widely shared NIA change. The changes with broader range across NIA are examined in §7.4.4.

7.4.1 Sequencing the 'central' KRDS Propagation Networks

The most complex array of isogloss boundaries within the KRDS area is found in a geographical 'corridor' approximately 50 kilometres wide which separates the western and 'central' KRDS speech communities.[13] This area of criss-crossing boundaries of propagation events is termed here as KRDS's 'western corridor of change'. The main phonological and morphological innovations whose ranges of propagation falter within this corridor are given in Figure 7.6.[14] The isogloss boundaries are labelled, and the key explains whether the innovation is found to the east (marked by 'e') or to the west (marked by 'w') of the boundary line. A summary of the linguistic character of each innovation is also given; the details are found in earlier chapters.

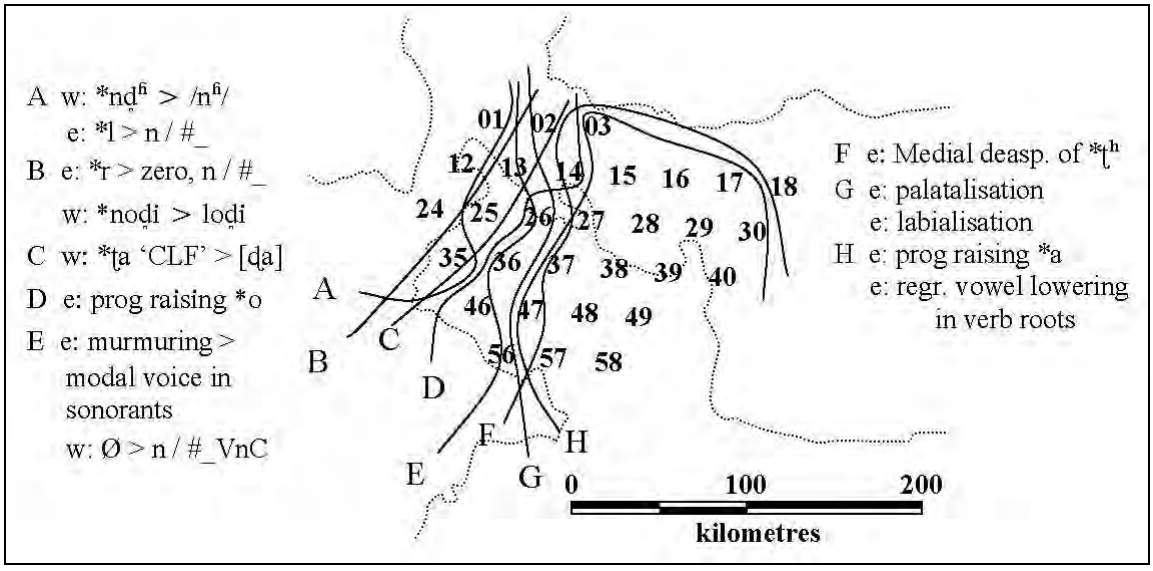


Figure 7.6: Isogloss boundaries in KRDS's western corridor of change

---

13 The inverted commas will not be repeated throughout, but are used here to draw attention to the intended meaning of 'central' as *geographically medial*, and not *more important* than the others. An alternative term could be 'midlands KRDS', analogous to the use of 'midlands NIA' rather than 'central NIA' to refer to Hindi etc., compare for example Masica (1991).

14 For more detailed analysis of the individual isoglosses see further below; for the data on which these isoglosses are based, see Appendix E of Toulmin (2006); for the experimental design and method used in data collection, see Chapter 2.

At this point in the historical reconstruction there are two methodological alternatives open to us. Confronted by such a complex dialectological pattern of innovations we might determine this to be the limit beyond which linguistic history cannot be reconstructed. After all, the relative chronology of these innovations cannot be established on linguistic grounds, because in general these changes do not bleed or feed each another (cf. §3.4.3.1), or on textual grounds, because in general these lects lack a written record which might guide the reconstruction of chronology (cf. §3.4.3.2). If this is the limit beyond which historical linguistic reconstruction is impossible, then it makes sense to either: (a) focus exclusively on the history of written languages, which afford us more opportunities to disambiguate chronologies of changes; or (b) conclude the study with a dialectological map, rather than an historical account—with a wave model, rather than a tree model. This statement summarises the present position of historical linguistic research in Indo-Aryan. The major studies (e.g. Chatterji 1926; Kakati 1962) focus on the lects with a tradition of written literature, the historical documents providing evidence (though not without some problems) as to the chronology of changes in spoken vernaculars over the centuries. Maniruzzaman (1977), who, on the other hand, departs from the focus on written lects to reconstruct the history of 'five dialects of Bengali', concludes without an historical account of the chronology of changes, but with a map of isogloss boundaries.

I have suggested in this study that in addition to linguistic seriation and textual sequencing of changes, there may be *sociohistorical* (that is, *historical sociolinguistic*) grounds for disambiguating the chronology of linguistic changes. Step V of the sociohistorical framework for historical linguistic reconstruction developed in §3.4.3.3 is reproduced below so as to guide the sequencing of changes in KRDS's western corridor:

V. Consider (i) the possible permutations of SCEs (divisions and integrations) which would account for the disjunction in PNs, (ii) the relative sociohistorical plausibility of each possible permutation, and (iii) the relative sociohistorical plausibility of a SCE as axgainst the co-existence of the PNs within a complex SC. Accordingly, reconstruct the chronology of PEs by selecting the most plausible sociohistorical explanation.

It has been argued in §7.3.1 above that the old (or proto) Kamta period of 1250–1550 was closed, and the middle KRDS period of 1550–1787 inaugurated, by division of the SC in conjunction with sociopolitical and linguistic expansion. Following this division, three possibilities come into play which must be compared and evaluated for sociohistorical plausibility: (1) SC division, (2) SC (re)integration, and (3) coexisting PNs within a complex SC. If the linguistic history was characterised primarily by *divisions* after the proto Kamta period, then the innovation with the greatest range would have been the 'first cab off the rank', followed by other changes in order of decreasing range. The hypothesis of *division* entails sequencing isoglosses A and B as historically prior to isoglosses G and H, in Figure 7.6. If, however, the linguistic history was characterised primarily by *reintegration* of partially differentiated SCs (after the initial division of the proto Kamta SC and its language), then isoglosses G and H would have preceded A and B. The third possibility is that the SC that developed after the division of proto Kamta was so complex in its network structure that it sustained contemporaneous networks with ranges A, B, G, H and so on.

I will argue below that the old course of the river Tista shaped the division of proto Kamta into central and western PNs, but that a later change to the course of this river—coupled with district reorganisation—led to a reintegration between portions of central and western

KRDS. This account of division-followed-by-partial-reintegration (second option above) is more plausible than the other two sociohistorical possibilities given what we know of the social history of the area. The problem with SC division-followed-by-further-division (first option above) as a sociohistorical explanation is that, prior to the shift of the river's course in 1787, there is no record of sociohistorical conditions which would account for isoglosses A and B as Propagation Networks. However, such sociohistorical conditions emerged *after the river shifted* due to the sociopolitical reorganisation that occurred during the colonial period. The third sociohistorical possibility, of a complex SC accounting simultaneously for all the PNs in the western corridor of change, is also problematic because the old course of the river (which most probably conditioned isoglosses G and H) only overlapped with the reorganised district boundaries for a decade. It is unlikely that all these isoglosses are the result of propagation during a single decade. Much more likely is the hypothesis that the isoglosses reflecting colonially-reorganised districts reflect a Propagation Network of interaction which *replaced* the older PN after the river changed course.

In order to unpack this sociohistorical argument, §7.4.1.1 explores the sociohistorical conditioning for the eastern edge of KRDS's western corridor of change, and the same is done for the western edge in §7.4.1.2.

7.4.1.1 The eastern limit of KRDS's western corridor of change

The eastern limit of this corridor of change may be defined by isoglosses F, G and H. The fuller representation of the dialect data for each of these changes is given in Figures 7.7 to 7.11. (The full data set are in Appendix D of Toulmin 2006). Dark shading in the figures below indicates the categorical, or near-categorical presence of the change in the data collected at the site in question—that is, there was little or no variability found with respect to this feature during the interviews with speakers at this site. Light shading (as in Figure 7.8) indicates that the change is variable in the data for that site.

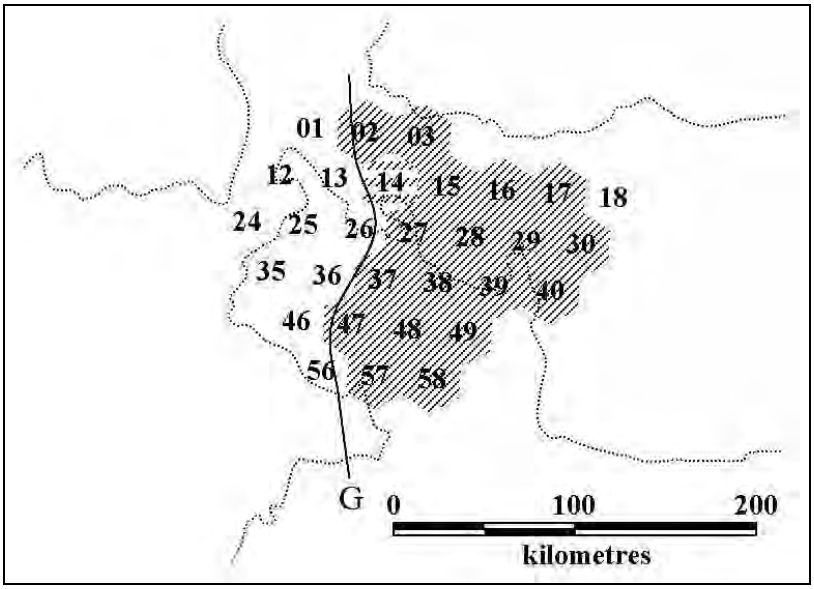


Figure 7.7: Dialect geography of [PI 6.] (palatalisation [i_a])[15]

[15] See items 2, 3, 4, 5 and 17 in Appendix E of Toulmin (2006).

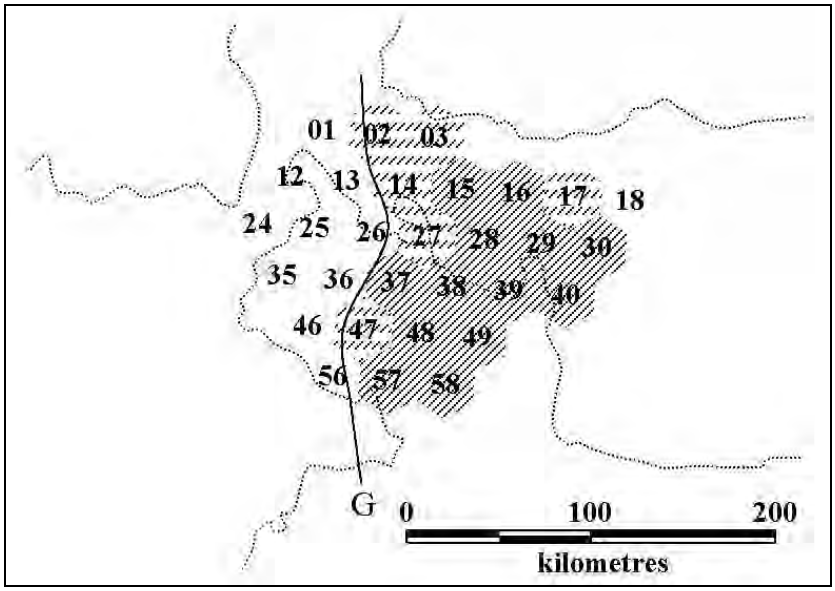


Figure 7.8: Dialect geography of [PI 7.] (labialisation/palatalisation [u_a])[16]

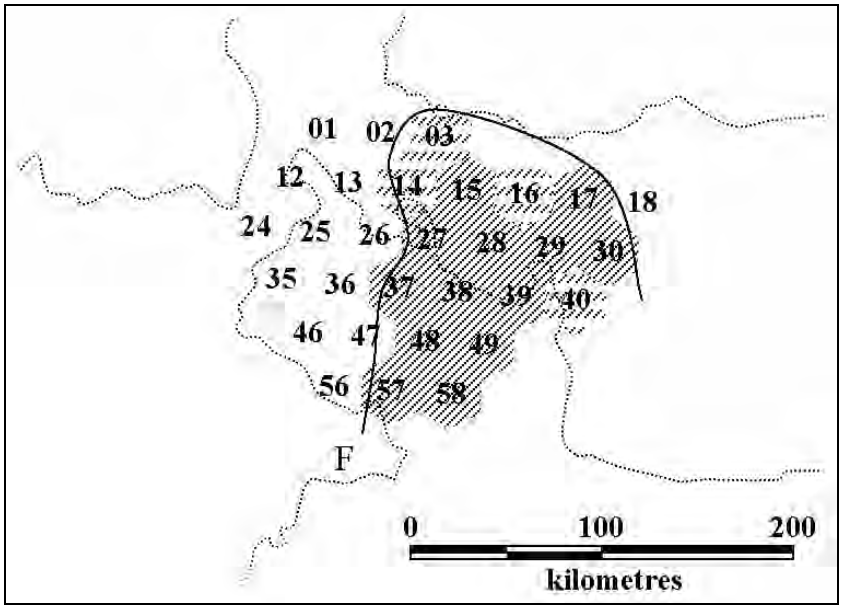


Figure 7.9: Dialect geography of [PI 10.] (medial deasp. of *t^h)[17]

[16] See items 1, 7, 10 and 22 in Appendix E of Toulmin (2006).
[17] See items 11, 12, 17 and 25 in Appendix E of Toulmin (2006).

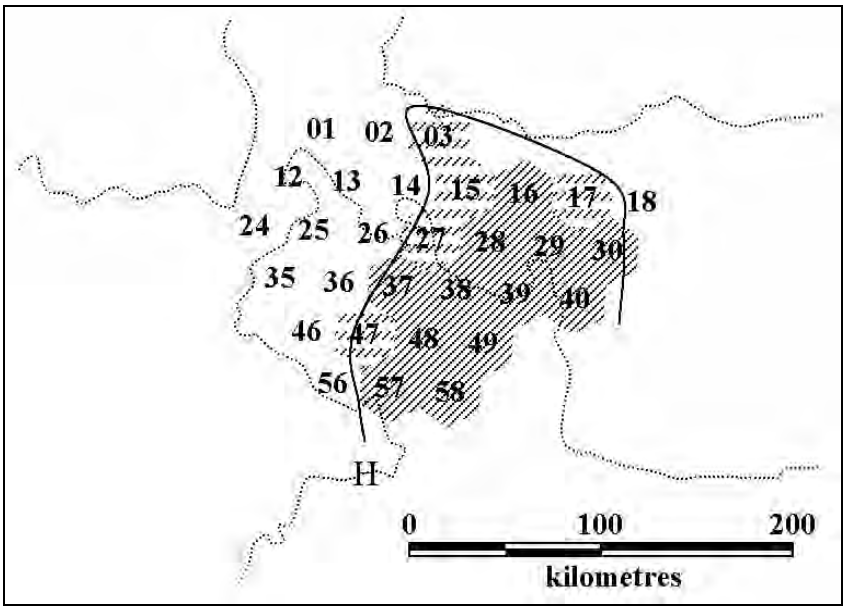


Figure 7.10: Dialect geography of [PI 24.] (progressive raising of *a)[18]

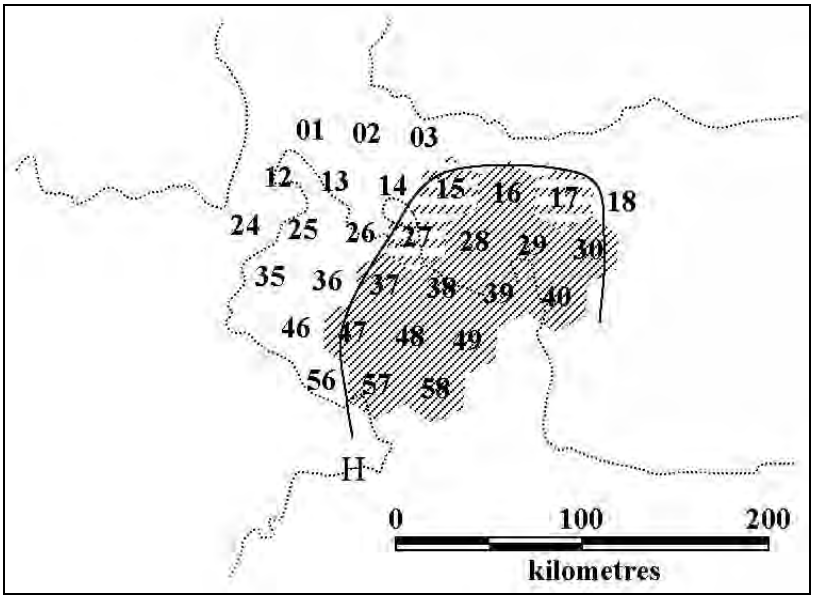


Figure 7.11: Dialect geography of [PI 26.] (regr. vowel lowering in verb roots)[19]

The geographical extent of propagation of these innovations along the western corridor does not correspond with contemporary sociopolitical, geographical or social boundaries, but with the *old course of the Tista river*. This old river ran from north to south, and was, until around AD 1787, the most marked geographical phenomenon, and the most enduring political boundary, of the KRDS area. Majumdar and Sarkar (1943) and others, observing that the Tista ran due south from Jalpaiguri in three streams—the Karatoya, the Atrai and

[18] See items 1, 2, 3, 4, 5, 7, 10, 17 and 22 in Appendix E of Toulmin (2006).
[19] See items 35, 37 and 38 in Appendix E of Toulmin (2006).

the Purnabhaha—derive the name Tista from tri-srota 'three streams' in Sanskrit. Clark (1969:98) explains the relation between the old Tista and Karatoya rivers:

> The history of the paradelta shows that the Karatoyā and the Tista have been closely associated. The Tista flowed into the Karatoyā, giving that river much of its size and power during the years it served as an ethnic and political boundary. Since it carries the run off from the high rainfall (120–150 inches yearly) Sikkim Himalayas, it has always been as Spate points out, an 'exceptionally violent' river, and has frequent devastating floods during the monsoon season.[20]

As mentioned by Clarke, not only was the river a major geographical phenomenon, it also functioned as an ethnic and political boundary at several points in the history of the region. During early medieval times, this river formed the boundary between the kingdoms of Kamrupa and Gauda. It also formed the boundary between the Koch and Gaur kingdoms before the expansion made by Biswa Singha's sons (Whyte 2002:25). After the fragmentation of the greater Koch kingdom, and for a time, this river separated the Koch kingdom from Mughal Bengal (Nathan 1936:804; cf. S.N. Bhattacharya 1943:241).

In contrast with its historical importance, the Atrai river is today but a small stream which nonetheless preserves the channel through which a major regional river used to flow. Test sites 26 and 47, shown in the figures above, are located within a few kilometres of that historical river course. Isogloss boundary G runs right along this old course from north to south; boundaries F and H also run along the same course, dividing sites 56, 36 and 26 from sites 57, 37 and 27, and slightly less precisely dividing between sites 46 and 48. Boundaries F and H differ from boundary G principally by not extending all the way north along the river course to sites 02, 03 and 14. The north-south course of the Tista river system is shown by the maps in Figure 7.1 and Figure 7.2, reproduced from Whyte (2002), and is also included in the maps in Ahmed (1936).

As reported by the Gazetteers, large earthquakes and floods during AD 1787 had a major impact on the course and unity of the Tista river system, effectively splitting a major river into several smaller rivers.

> Before 1787, the Atrai was one of the great rivers of North Bengal for through this channel the Tista used to discharge its water into the Padma [Ganga]. But in 1787, a great flood took place and changes occurred in the river system of this region. This was also due to earthquakes and earth movement. As a result, the Tista broke away from its old channel and found a new and capacious channel south-eastward and joined the Brahmaputra (Jamuna) … Since then the Atrai has [lost] its former importance, but [is] still navigable by large country boats during the rainy season. (Bangladesh District Gazetteers, Rajshahi, 1976:3, cited in Islam 1992:7)

This event had a catastrophic effect on the lives of the inhabitants of the low lying flood-plains. Clarke (1969) cites Henry Frowde's statement in the Imperial Gazetteer of India that *one sixth of the local population died in the disaster*.

On the one hand then, we have the geographical course of the River Tista-Karatoya-Atrai which was a pre-modern political boundary and ran north to south until AD 1787. On the other hand, we have the dialect geography of several innovations which (1) are diagnostic of PEs, and (2) share substantial portions of their western boundary with this old course of River Tista. Based on these correspondences in geography, I propose that the river Tista was, before its division, a sufficient boundary to interaction between speakers

---

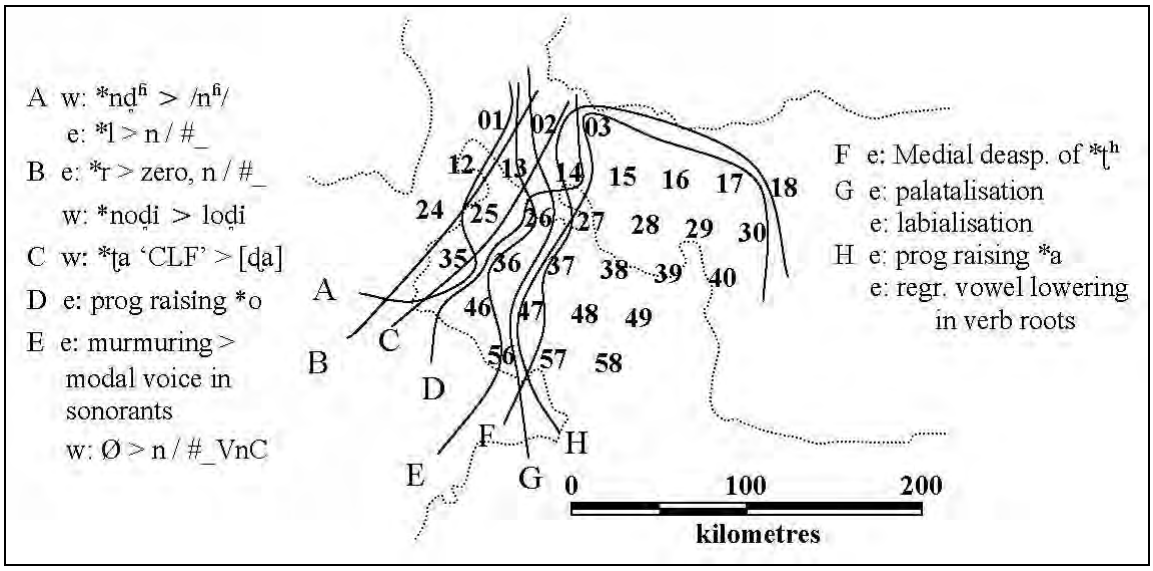
[20] The reference is to Spate and Farmer (1954).

living on either side that local changes propagated among speakers on one side were not adopted by speakers on the other side. It may be mentioned in support of this argument that (1) the rivers of today's north Bengal are in general fast running, and among them the Tista is at times 'exceptionally violent' (see quote from Clarke above); (2) the previous channel of River Tista was considerably wider than any of the present-day rivers of north Bengal (not including the Brahmaputra in the east); and (3) this major river also functioned as a sociopolitical boundary at several points in the history of the region before its shift in course.

A causal connection between (a) the zone of interaction bounded by the old course of River Tista and (b) the propagation of innovations shown in Figures 7.7 to 7.11, provides a *terminus ante quem* for the associated innovations. The shift in Tista's course would have resulted in a major restructuring of patterns of social interaction across North Bengal and the Koch Behar kingdom. For the old course of Tista to have so precisely conditioned the extent of propagation of linguistic changes, *those changes must have been propagated prior to the change in the river's course*. The easternmost limit of KRDS's western corridor of change is accordingly dated as prior to AD 1787 (when the river shifted).

7.4.1.2 The western limit of KRDS's western corridor of change

The western limit of the corridor runs not in a north-south direction but from north-east to south-west, and may be defined by the isoglosses A, B and C of Figure 7.6, repeated here for ease of reference.



Reproduction of **Figure 7.6:** Isogloss boundaries in KRDS's western corridor of change

Fuller representation of the ranges of these PNs is given in Figures 7.12 to 7.16. The solid shading in the figures below (which contrasts with diagonal shading) indicates the presence of the innovation in the wordlist data collected at RL, KS and MH during the first stage of fieldwork (cf. Chapter 2).

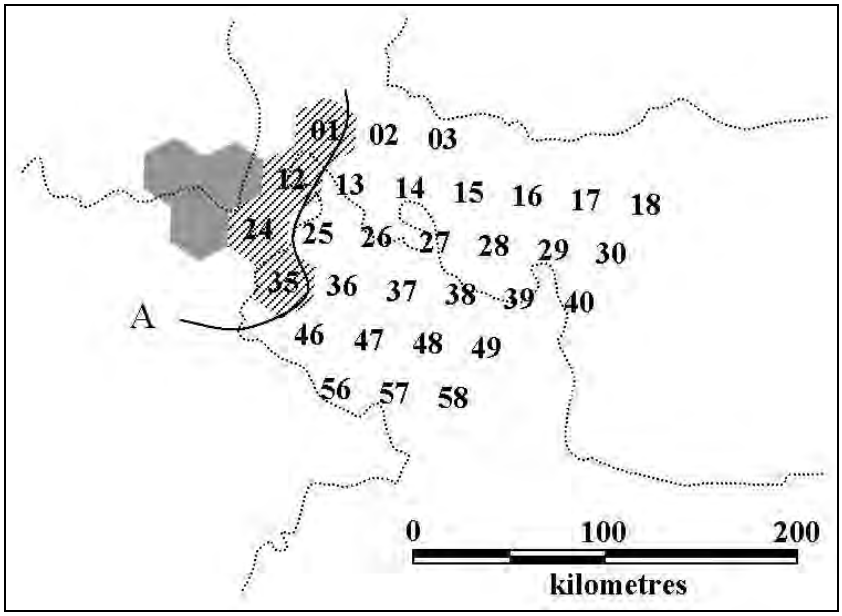


Figure 7.12: Dialect geography of [PI 17.] (*nɖʱ > /nʱ/)[21]

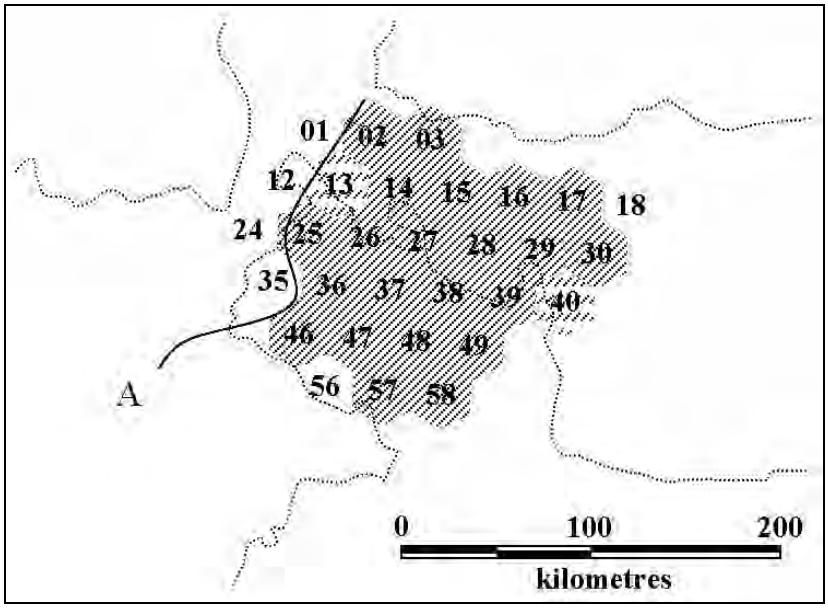


Figure 7.13: Dialect geography of [PI 14.] (initial *l > /n/)[22]

[21] See item 37 in Appendix E. This change is supportive, but not diagnostic of a PE (cf. 4.3.12).

[22] See item 12 in Appendix E. The innovation [PI 14.] is not a diagnostic change (cf. 4.3.11), but as the range is almost identical as that of [PI 15.] (which is diagnostic), it is likely that they resulted from a single propagation event.

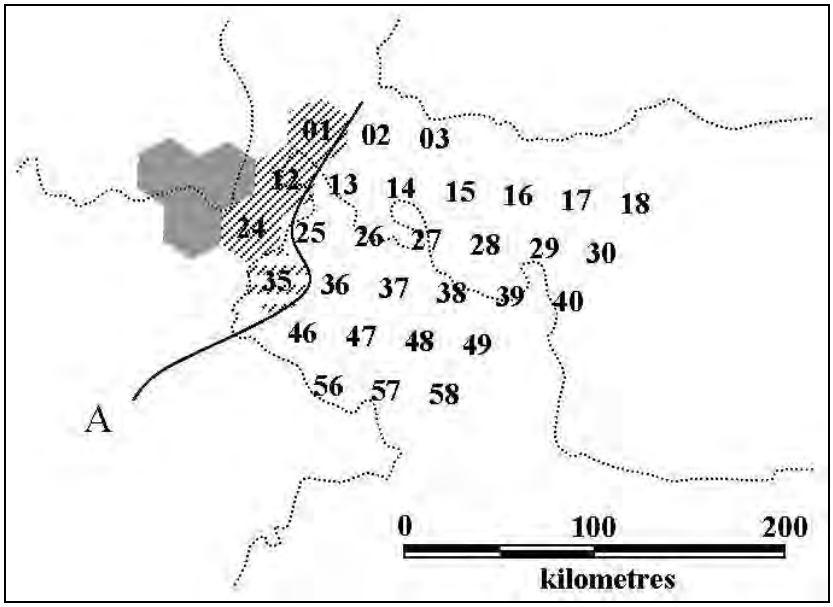


Figure 7.14: Dialect geography of initial *n > /l/[23]

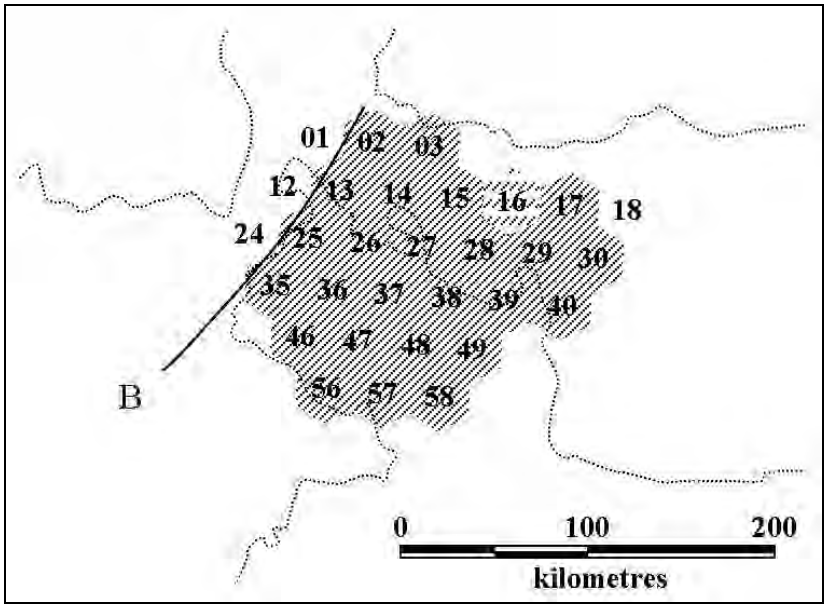


Figure 7.15: Dialect geography of [PI 15.] (initial *r > Ø)[24]

[23] See item 13 in Appendix E.

[24] See item 14 *[rɔʃuɳɔ, rɛʃuɳɔ, lɛʃuɳɔ] 'garlic' in Appendix E, which is a Tadbhava word. The variant collected at site 16 is /neʃun/ which seems to be derived from *lɛʃuɳɔ rather than *rɛʃuɳɔ. Hence, neither loss nor retention of initial *r is attested by this item. The hexagon is shaded in Figure 7.15 because the wordlist data collected within a few kilometres of site 16 shows variable loss of initial *r.

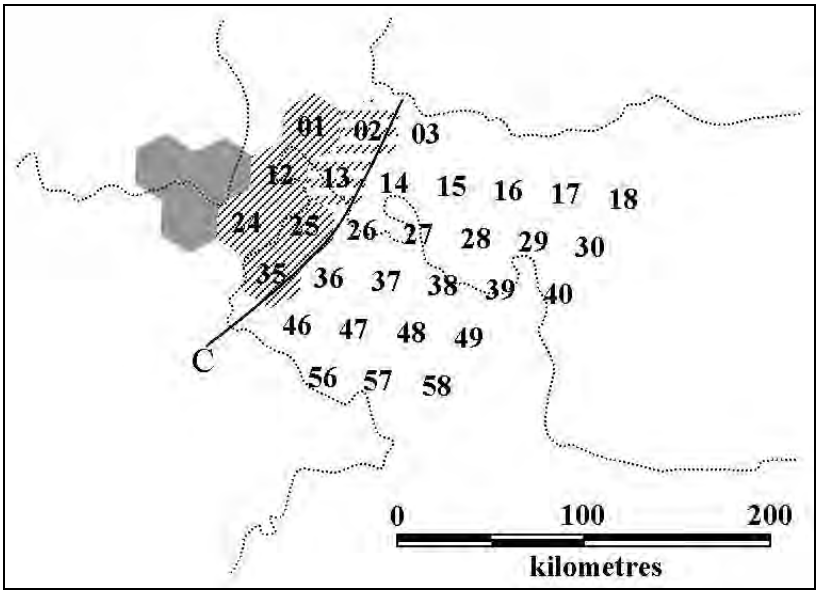


Figure 7.16: Dialect geography of *ʈa 'CLF' > [ɖa] / C[sonorant] _ [25]

The range of propagation of these innovations does not correspond with major geographical features, but rather with the major sociopolitical boundaries established by the British after they secured control of the region in AD 1773. Notice that the ranges of PNs shown in Figures 7.13 to 7.17 divide sites on the north-west (1, 12 and 24; also 35, except for PI 15) from sites to the east and south. The region was divided along these lines—between Bihar in the north-west, and Dinajpur and Rangpur in the east and south—only since the time of British rule (AD 1773). The district of Dinajpur (Dinagepour) is shown in James Rennell's 1781 Bengal Atlas with these boundaries (see Rennell and Ambashthya 1975). The western limit of Dinajpur at that time followed a small river called the Nagar (Nagore) river, and today this same river forms the international border between West Bengal and Bangladesh (marked in Figures 7.12 to 7.16).

Prior to British rule, the river Nagar had no significance as a boundary between administrative divisions—on the contrary, the river Nagar flowed directly through the centre of the Sarkar (Mughal equivalent of a district) called Tajpur (see Habib 1982: plate 11). The western boundary of Tajpur Sarkar was the Mahananda river, which lies 30–40 kilometres to the west of the isoglosses shown in Figures 7.12 to 7.16. The Mahananda river runs south from the eastern border of Nepal, shown in those same figures. The eastern boundary of Tajpur Sarkar lay between the rivers Nagar and (the much greater) Tista. In short, the western limit of KRDS's western corridor of change corresponds very poorly with the Mughal administrative divisions, and much more neatly with the British reorganisation of districts.

On the basis of these geographical observations, I propose a causal correspondence between the borders of district organisation established by the British and the extent of propagation shown in Figures 7.12 to 7.16. The same pattern of propagation was not found in the set of changes discussed earlier (Figures 7.7 to 7.11), plausibly because the two patterns reflect historically distinct networks of speaker organisation and interaction. The historical changes which best account for this change in speaker organisation and

[25] See item 53 in Appendix E.

interaction occurred at the end of the 18th century when: (1) the River Tista shifted its course, and (2) the British administration restructured the district boundaries. These restructured district boundaries provide a plausible sociohistorical cause for the western limit of the western corridor of change.

However, the impact of this administrative reorganisation on patterns of social interaction and linguistic performance would hardly have been instantaneous. It probably took a few generations for the new sociopolitical boundaries to stamp their mark on patterns of interaction and linguistic propagation. We may assumed though, that the impact on speaker interaction of the administrative reorganisation would have been exacerbated by the catastrophic events of AD 1787 (cf. §7.4.1.1). An earthquake and flooding which killed one sixth of the local population would have caused major destruction of villages and led to considerable rebuilding of lives within the region. The rebuilding would have been both of physical things like houses and farms, but also of social things such as family relations. It is plausible that this rebuilding phase sped up the process of reshaping patterns of social interaction (and hence propagation events) along the new district lines.

7.4.1.3 Results of the sociohistorical sequencing of 'central' KRDS PNs

Recall the shape of the argument for sociohistorical sequencing developed in §3.4.3.3:

If PNs are reconstructed with a disjunction in their ranges,

And a SCE is, on balance, more sociohistorically plausible than the co-existence of these PNs within a complex SC,

And a particular directionality of SCE (i.e. either SC division or integration) is more plausible for sociohistorical reasons,

Then this plausible directionality of the SCE also supports a particular sequencing of the PEs.

In the case of KRDS's western corridor of change, there is a disjunction in ranges, and I have argued that there are sociohistorical reasons for considering a particular directionality of SCE to be more plausible than the alternatives. The SCE is as follows: (a) earlier division along the old Tista river course, with (b) subsequent reintegration between communities on either side after the river shifted course (the sociolinguistic reintegration also aided by the sociopolitical integration brought into effect by the colonial powers at about the same time). This particular directionality of SCEs is more plausible than the alternative scenarios given what we know of the social history of the area. This directionality of Speech Community Events entails a particular chronological relation between the linguistic Propagation Events: *PEs associated with the old river course preceded PEs associated with the reorganised districts*. The changes that preceded the shift in river course in AD 1787 must also have occurred subsequent to the division of the proto Kamta SC in AD 1550 (following the expansion of the Koch kingdom). This period of time, delimited for us by expansion at one end and earthquake at the other, is defined in this study as the 'middle KRDS' stage of linguistic history. From this point onwards it becomes historically appropriate to talk of Kamta/Rajbanshi/Deshi Bhasha, rather than just 'Kamta' because it is during this period of AD 1550–1787 that the term Rajbanshi is first attested in historical documents, and it is during this period (under the Muslim rulers) that the term Bengal is increasingly used to refer to the regions earlier termed Gauda, Banga, and so on.

7.4.2 The westwards migration of proto Kamta speakers

The sociohistorical arguments above suggest that speakers of the proto Kamta language migrated westwards—across the Tista-Karatoya—in conjunction with the expansion of the Koch kingdom. The communities of emigrants took with them the innovative proto Kamta features—but having distanced themselves from the central KRDS SC, they did not participate in the propagation of common innovations with the central SC during the period AD 1550–1787. It was during this period that changes such as the palatalisation and labialisation of stops ([PI 6.] and [PI 7.]) plausibly occurred.

While several changes have been shown above to line up with the eastern bank of the old course of Tista, the dialectological data collected for this study provide less data on changes that might line up along the western side of the river. The comparative reconstruction has uncovered four possible candidates for 'old west bank' changes. They are:

[PI 18.] *ɛnd̪urɔ > /nid̪ur/ 'rat' {KS, RL, MH, TH}. Supportive, not diagnostic.

[PI 36.] *ɔw > /u/ {KS, RL, MH, TH}. Supportive, not diagnostic.

[PI 25.] *o > /ɔ/ / _Ca {KS, RL, MH, TH}. Diagnostic.

[MI 38.] *-rɔkɔm > *-rɔŋ 'like, similar to' {KS, RL, MH, TH (Hindus, not Muslims)}. Diagnostic.

Dialectological data were collected for the first and second, but not the third and fourth of these changes. The latter changes are sequenced similarly to the first two on the basis of economy of reconstruction, but this sequencing may need to be revised if and when the relevant dialectological data become available.

The dialectological data for [PI 18.] are represented in Figure 7.17. The solid shading once again indicates the presence of the innovation in the wordlist data collected at RL, KS and MH during the first stage of fieldwork.

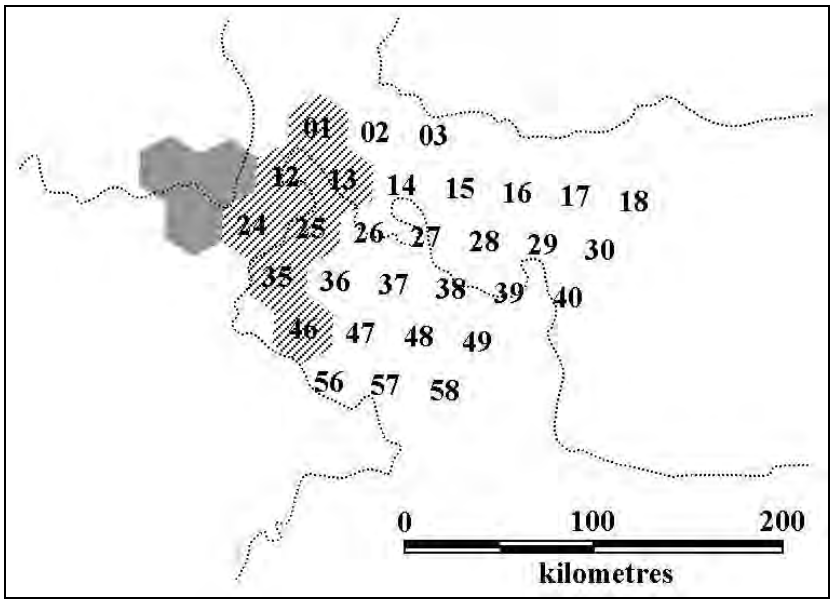


Figure 7.17: Dialect geography of *ɛnd̪urɔ > /nid̪ur/ 'rat' [26]

[26] See item 32 in Appendix E.

The correspondence between the range of [PI 18.] and the old course of Tista is suggestive, even if not quite as exact as was seen above for the central KRDS changes. It is possible that [PI 18.] may have been present at sites 36 and 56 before the reintegration with central KRDS but subsequently lost due to borrowing of /ɛnd̪ur/ 'rat'.

The second possible proto-western KRDS change for which data was collected is [PI 36.], demonstrated by the divergence in primary agreement ending for first person singular subject. Generally, east of old Tista's course the ending is /-o˜; -o/, while west of the old course the ending is generally /-u˜; -u/. A more precise description of the dialectological range of the two variants is given in Figure 7.18.

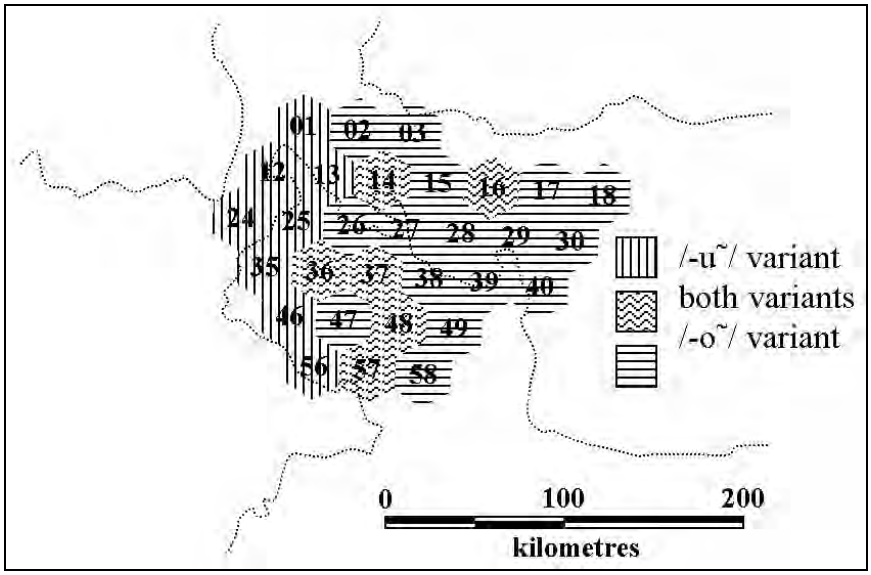


Figure 7.18: Dialect geography of /u˜/ vs. /o˜/ '1SG' in AGR.I[27]

This difference in personal endings is explained in §6.4.1.1 as the result of different reflexes of the inherited phonological sequence *-ɔw˜. East of old Tista, the reflex is /-o˜/; west of old Tista the reflex is /-u˜/. The irregularity in Figure 7.18 on the eastern side of Tista (with instances of /-u/ at sites 14, 16, 37, 48 and 57) could be the result of *analogical extension* of /-u/ '1SG' from the past tense system (AGR.IIA) to the primary system of endings (compare the agreement systems of TH and RP in §6.3.4 and §6.3.6). The irregularity on the western side of Tista (with /-o/ collected at sites 26 and 36) could be *borrowing* as a result of exposure to the Tista-east norms after the shift in river course and the SC reintegration (similar to [PI 18.] above).

Together, these two changes ([PI 18.] and [PI 36.]) provide suggestive, though not conclusive evidence for a Propagation Network on the western bank of old Tista during the Middle KRDS period. Dialectological data for [PI 25.] and [MI 38.] may render the hypothesis more robust. At the present time, the reconstruction remains in some doubt and thus proto western KRDS is prefixed with a question mark in the tree diagrams of KRDS linguistic history: '?p-wKRDS' (see Figure 7.5).

More localised innovations west of the old river course are not hard to find, and will be discussed in §7.5.2. A chronology for these more localised changes earlier than AD 1787 cannot be proven at this stage.

---

[27] See items 83 and 65 in Appendix E.

7.4.3 The formation of an eastern KRDS speech community

The north-eastern-most extent of the central KRDS changes is between sites 17 and 30 (Gosaigaon and Bogribari) and site 18 (near Kokrajhar) in the figures above. *The boundary is the same for PEs dated as Middle KRDS changes and for PEs dated as Modern KRDS changes* (cf. §7.5). That is, the KRDS 'eastern corrider of change'—separating central and eastern SCs—seems to have been stable since AD 1550 between Kokrajhar and Gosaigaon or Bogribari. This dialectological boundary is supported by the large wordlist data collected at BN (just east of site 18) and BH (near site 16). The sociohistorical conditioning of this sociohistorical division is not clear (cf. §7.5.4).

7.4.4 Partial reintegration between KRDS and other NIA SCs

There are several innovations that are plausibly assigned to the middle KRDS period, and which are shared further afield than just KRDS:

[PI 9.] *C[+breathy voice, +continuant] > [+modal voice] / V_V {SH, RP, BH, BN, Oriya, SCA, SCB} ([tentatively] after C16th, after rhoticisation). Diagnostic.

[PI 10.] Loss of aspiration in all intervocalic consonants {RP, BH, ?SCB, ?Oriya}. Diagnostic.

[PI 12.] *ɳ, *ɭ > /n, l/ {KRDS, SCB, SCA, Maithili, Hindi, etc.} (C15th or later). Possibly diagnostic based on sociohistorical plausibility (cf. §7.4.4), but broader NIA reconstruction is necessary in order to verify the diagnostic value of this change.

[PI 26.] *i, *u > /e/, /o/ / #(C)_C-V[- H] (verb root) {RP, BH, SCB}. Diagnostic.

[PI 30.] *$V_1C_X^i$ > *$V_1^iC_X$ /-C_YV {irregularly in KRDS, Middle Bangla, Oriya of C15th, Damrupi, Asamiya, …}. Diagnostic value unclear.

[PI 38.] *ɔ lost word finally {KRDS, SCB, SCA, Hindi, Bhojpuri etc.} (chronology uncertain).

The loss of final *ɔ [PI 38.] occurs in all KRDS lects, as well as in Bangla, Asamiya, Hindi, Magahi, Nepali, and many other NIA lects besides. Final *ɔ is maintained in Oriya and a small number of other NIA lects (cf. Masica 1991:196). Somewhat similarly, the dental and postalveolar nasals and laterals have undergone merger in a number of NIA lects, notably Bangla, Asamiya, Hindi, and the Bihari lects. Oriya is exempt from this change, as are the western NIA lects (e.g. Gujarati, Punjabi, etc.). There is a considerable overlap between the ranges of propagation of these changes, at least in eastern and midlands NIA. It is not out of the question that these changes were propagated *at the same time,* and through the same networks of speaker interaction. In this respect, we might consider the early 16th century expansion of the Mughal Empire as a possible cause for increased interaction between the Indo-Aryan midlands and eastern regions. Sarkar (1943:216) writes:

> The period of Mughal imperial rule over Bengal witnessed the working of certain new forces which have completely transformed Bengali life and thought and whose influence is still operating in the province. In one word, during the first century of Mughal rule (AD 1575–1675), the outer world came to Bengal and Bengal went out of herself to the outer world, and the economic, social and cultural changes that grew out of this mingling of peoples mark a most important and distinct stage in the evolution of modern Bengal. Indeed, there has been nothing in our province's past history at all comparable to it except the modernisation which we owe to the British influence.

The hypothesis of a sociohistorical connection between the Mughal expansion and the propagation of [PI 12.], [PI 30.] and [PI 38.] can only be properly verified in the context of a much broader reconstruction of NIA linguistic history.

The next change to be considered is the change from murmured to modal voicing in sonorants ([PI 9.], cf. §4.3.4). This change occurred in Oriya, Asamiya and Bangla, though not in the neighbouring Bihari lects. Within KRDS, the range of the innovation is represented by isogloss E in Figure 7.6 above. The boundary of the isogloss runs through the middle of KRDS's western corridor of change, which makes assigning chronology on sociohistorical grounds difficult. Textual evidence from neighbouring standardised lects suggests that this change occurred during the middle KRDS period. The changes [PI 10.] (general loss of intervocalic aspiration) and [PI 26.] (regressive lowering in verb roots) have been reconstructed for the same period, based on the alignment of their range with the old course of Tista. Both [PI 10.] and [PI 26.] are shared by cKRDS and Bangla, and the former of the two changes (deaspiration) may also be common with Oriya (cf. P.C. Majumdar 1970:xxxiii).

We turn now to considering the sociohistorical phenomena attested for the middle KRDS period which might have facilitated the propagation of these changes between KRDS and Bangla in the case of [PI 10.] and [PI 26.], and further up the Brahmaputra valley to Asamiya in the case of [PI 9.].

As described in §7.3.1, the 16th century saw the rapid expansion of the Kamta kingdom under Nara Narayan's reign through the initiatives of General Sukladhvaj. However, before the close of the same century, the kingdom had begun to shrink in size, and a rift within the ruling family of Kamta kingdom saw the kingdom divided along the Sankosh river (Gait 1905). Wars between the sister kingdoms led to Koch-Kamta's dependence on the Mughal armies coming from Bengal to defeat the rebel eastern Koch kingdom. A Mughal presence under the administration of Bengal was established in the region geographically between the Koch and Ahom kingdoms (Bhattacharya 1929, cited in Whyte 2002:28). A temporary weakening of the Mughal kingdom in the mid 17th century led to joint efforts by the Koches and Ahoms to expel the Mughal presence, and the Koch armies campaigned 'possibly as far south as Dhaka' (Whyte 2002:28). The Mughal powers responded with a massive campaign under Mir Jumla, who marched his armies north from Dhaka and established fleeting victories over the Koch and Assam kingdoms, before disease and popular revolts forced the invaders' withdrawal. The geographical directions of these large scale movements of armed forces in the mid 17th century is sketched in Figure 7.19.

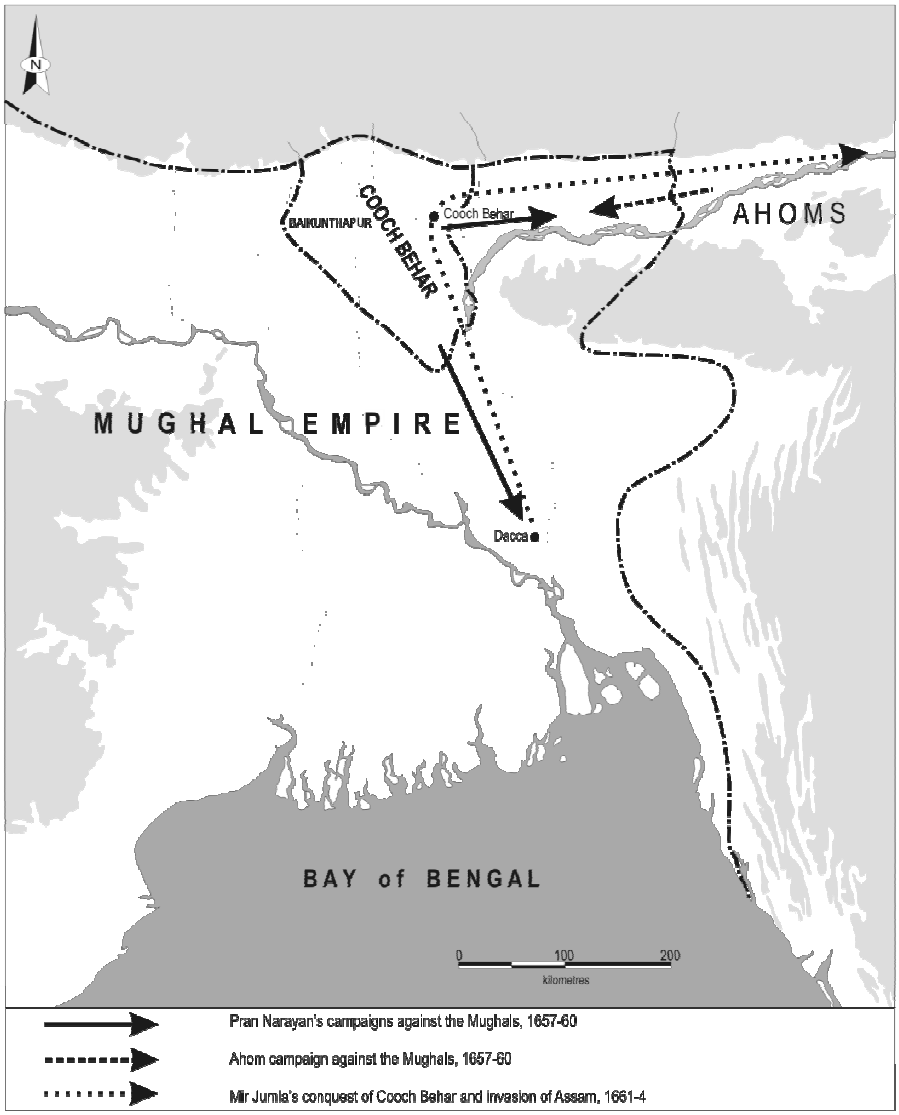


Figure 7.19: Pran Narayan's and Mir Jumla's campaigns c. 1660, reproduced from Whyte (2002:459)

Whyte goes on to say that as a result of the Mughal invasions in the late 17th century 'disbanded Mughal soldiers had occupied lands inside the remainder of Cooch Behar' (2002:31, citing D. Majumdar 1977). There is no doubt that this century was a tumultuous time for the inhabitants of the Cooch Behar kingdom, and also that it threw them into contact with speakers from south Bengal to an extent that had not happened in the preceding centuries. I propose therefore that it was during the 17th century that interaction with speakers from south Bengal led to the propagation of changes [PI 9.] (murmured > modal voicing), [PI 10.] (loss of medial aspiration), and [PI 26.] (regressive lowering) between the speakers of cKRDS and south Bangla lects.

This concludes the sociohistorical reconstruction of linguistic history for the middle KRDS period. The reconstruction to this point has been presented by way of the adjusted tree diagram in Figure 7.5 above.

7.5 The modern KRDS period: local innovations and the influence of standard languages over the last 220 years

The basic trend during the middle KRDS period (AD 1550–1787) was that local and wide-scope innovations were propagated somewhat concurrently. The same trend has continued into the modern period (after AD 1787) with the added effects of diglossia.

7.5.1 Diglossia during the Modern KRDS period

Diglossia is defined by Ferguson (2000:75) as:

> a relatively stable language situation in which, in addition to the primary dialects of the language (which may include a standard or regional standards), there is a very divergent, highly codified (often grammatically more complex) superposed variety, the vehicle of a large and respected body of written literature, either of an earlier period or in another speech community, which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any sector of the community for ordinary conversation.

The codification of NIA languages such as Bangla and Hindi, and their promulgation as superposed varieties, is usually attributed to the British period (Chatterji 1926:134; Kakati 1962:16–17; Masica 1991:29), though naturally these processes had their roots in earlier times. The key events which led to situations of diglossia were the establishment of Fort William College at Calcutta in AD 1800 and the launch of several printing presses in Calcutta during the first decade of that century. Chatterji writes regarding Standard Bangla that 'the advent of Western learning brought in a sudden demand for a prose style' (1926:134). A standard variety of Bangla began to be used for purposes of education, administration, and elite correspondence across the province of Bengal, including much of the KRDS area as well as today's Assam. The British colonisation had a similarly significant impact on the development of Modern Standard Hindi (MSH, alternatively termed Standard Colloquial Hindi, SCH), as Shapiro (2003:255) writes:

> To a great extent the emergence of MSH can be seen as a phenomenon that is thoroughly entwined with the sweeping political, social and communal changes that took place in North India between the establishment of the British Rāj in 1858 in the wake of the Great Rebellion of 1857-8 and the granting in 1947 of independence to India and Pakistan. The roots of these changes, of course, go back earlier.

In the case of Asamiya, a standard written variety also began to be promulgated during the 19th century through the printing press established by the American Baptist missionaries at Sibsagar. However, Goswami and Tamuli (2003:398) attribute the codification of written Asamiya to the earlier period of the *Buranjis* (17th to 19th centuries).

Therefore, with the possible exception of Asamiya, it seems justified to date as post AD 1800 those changes which are caused by diglossia. That is, they seem to have occurred during the Modern KRDS period. The changes which fit into this category are shown in Table 7.3 and include (i) morphological loans from the superposed variety in diglossia, and (ii) changes involving structural convergence (phonological or morphological) with this variety.

Table 7.3: Tableau showing changes resulting from diglossia

Innovations	Superposed variety	KS	RL	MH	TH	SH	RP	BH	BN
PI 29. MI 10.	Hindi								
PI 19. PI 34.	Hindi								
MI 8.	Hindi								
MI 50.	Hindi								
PI 28.	SCB								
MI 9.	SCB								
MI 54. MI 55. MI 58.	SCB								
MI 11. MI 12. MI 35. MI 72. MI 25. MI 28. MI 30. PI 11. MI 4.	SCA								

Across the KRDS area the superposed variety differs between Hindi (SCH), SCB and SCA. The different relations of diglossia occur in mutually exclusive regions: lects KS, RL and MH exist in diglossic relations with Hindi; lects TH, SH, RP and BH in relation to SCB; and the lect BN exists in a diglossic relation with SCA. The degree of influence which the superposed variety has exercised on each of these lects is not the same. BN is the lect most thoroughly influenced by a superposed variety (SCA), followed by KS and RL which are influenced by Hindi, and TH which is influenced by SCB.

Structural similarities resulting from diglossia are not diagnostic of PEs (cf. §3.4.1.2). They indicate the influence of a superposed lect upon the vernacular, not the propagation of a variant between the vernaculars of interacting speakers. Thus, the similarities shared by (for example) KS, RL and MH as the result of diglossia with Hindi do not make them a phylogenetic subgroup, because subgroups are defined by PEs (cf. §3.4.1.2). Accordingly, in the depiction of linguistic history, relations of diglossia are marked differently to phylogenetic relations. The latter are marked with *solid* horizontal double lines, as in Figures 7.4 and 7.5 above. Relations of diglossia are marked instead by *broken* double horizontal lines, see Figure 7.20. The broken *single* horizontal arrows connect KRDS lects with their respective superposed varieties. Thus, the meaning of horizontal arrows in Figure 7.20 is distinct from the meaning of vertical and diagonal arrows—the latter indicating inheritance (cf. §3.4.4).

It has not been possible, as yet, to collect dialectological data of equal detail for all the changes in Figure 7.20. Consequently, the chronology of some changes is not presently assignable by sociohistorical sequencing. For such changes, sequencing has been hypothesised using the principle of economy of reconstruction. (A hypothetical, unverified chronology is indicated by a prefixed question mark.) Sequencing is hypothesised for unassigned changes as follows: examining the distribution of the change among the eight sample KRDS lects, if there is a correlation with the distribution of other sequenced changes, then the same sequencing is hypothesised for the unassigned change as for the assigned change. These sequencing hypotheses should be reviewed once the relevant dialectological data become available.

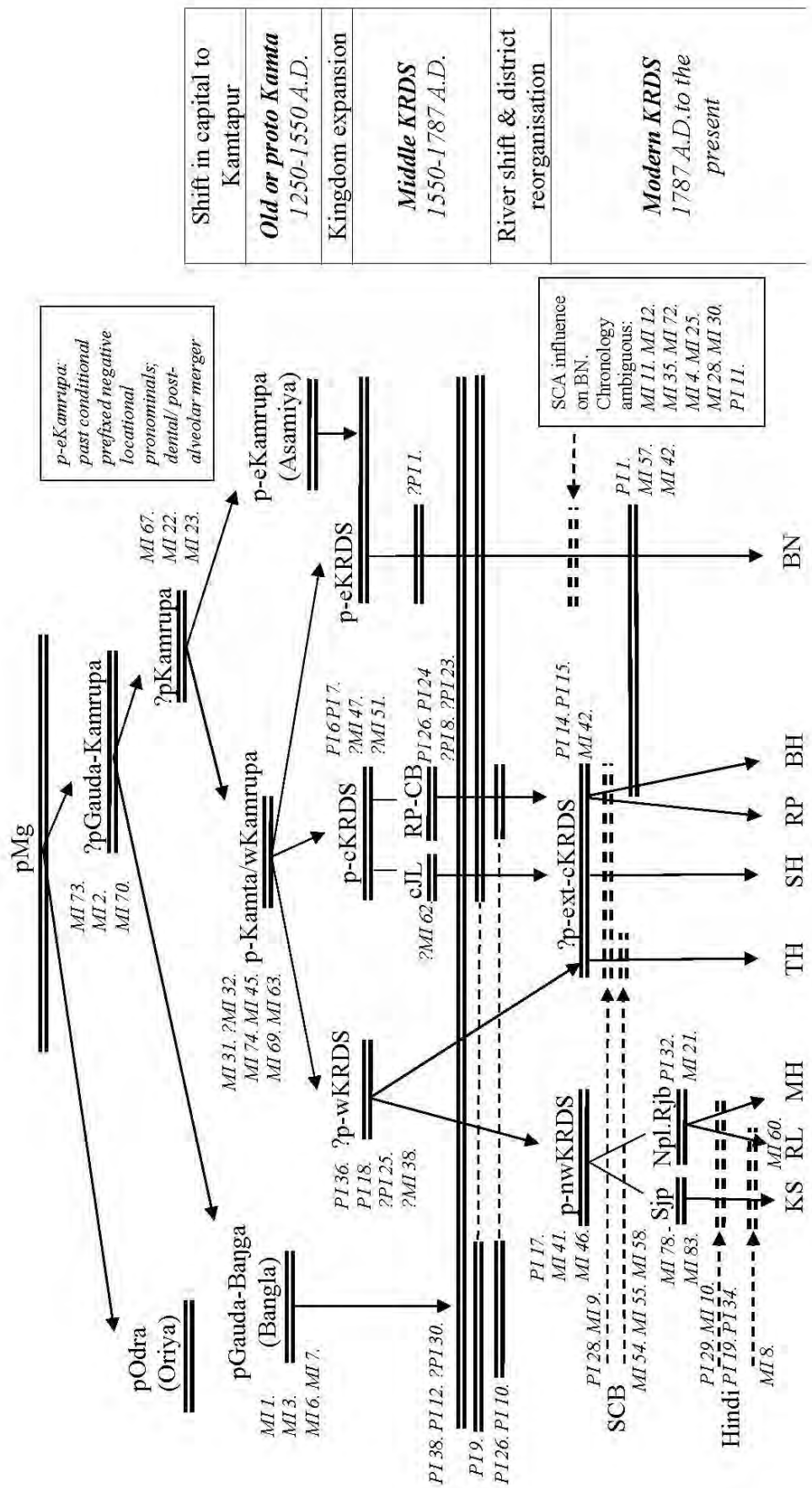


Figure 7.20: The linguistic history of KRDS from proto Magadhan, through proto Kamta, middle and modern KRDS, to the present

The historical relations depicted in the lowest portion of Figure 7.20 are discussed below.

7.5.2 North-west KRDS

The lects classed as 'north-west KRDS' are KS, RL and MH. These lects represent the 'Surjapuri' lect of Kishanganj district and the 'Rajbanshi' lect of Morang and Jhapa districts of Nepal. KS, RL and MH share certain innovations in common which define north-west KRDS as a Propagation Network and subgroup (§7.5.2.1). These three lects are also similarly characterised by diglossia with Hindi as the superposed variety (§7.5.2.2).

7.5.2.1 The north-west KRDS propagation network

There are three PEs identified in this study that diagnose a north-west KRDS PN. They are:

[PI 17.] Homorganic cluster of N C [+asp, +vc] > N [+asp] {irregularly in KS, RL, MH, TH}. Supportive, not diagnostic.

[MI 41.] *ɛlʱa > /alʱa/ 'now' {KS, RL, MH}. Diagnostic.

[MI 46.] *-ʈʰikna > /-ʈʰina/ 'place' {MH, RL, KS} (after [PI 20.] and [PI 30.]–[PI 33.]). Diagnostic.

Detailed dialectological data were not collected as part of this study across the whole of the north-west KRDS area.[28] However, the eastern limit of [PI 17.] has been depicted in Figure 7.12 and shown in §7.4.1.2 to be one of three changes whose limits of easward-propagation correspond to the administrative boundaries established by the British. Other changes with a similar range are: *n > /l/ word-initially in *nɔḍi 'river' (Figure 7.14); and *ʈa 'CLF' > [ɖa] after a sonorant consonant (Figure 7.16). The first of these two changes seems to be an irregular phonological change, given that in the phonological reconstruction of Chapter 4 there are only two etyma which have #n>l across all three of these lects (*nɔḍi 'river', id. #155; and *nabʱi 'navel', id. #349). The second change with this range is a morphologically-conditioned phonological change. Despite a lack of dialectological data for [MI 41.] and [MI 46.], these changes are sequenced similarly to [PI 17.] on the basis of economy of reconstruction. However, this sequencing may need to be revised once the relevant dialectological data become available.

[PI 17.] has already been reconstructed above as sociohistorically connected with the district boundaries between Purnia (Purneah), Dinajpur (Dinagepour) and Rangpur (Rungpour) established during the British period. Therefore, the proto nwKRDS Propagation Network is assigned to the Modern KRDS period (post-AD 1787).

Two more restricted PNs within nwKRDS are diagnosed by distinct sets of changes. The RL and MH lects are diagnosed as a subgroup by changes [PI 32.] and [MI 21.]. The KS lect is diagnosed as distinct from RL and MH by changes [MI 78.]–[MI 83.]. The RL-MH PN is labelled 'Npl-Rjb' (Nepali Rajbanshi), and the KS PN is labelled 'Sjp' (Surjapuri) in Figure 7.20. The relative chronology of these PNs with respect to proto north-western KRDS has not been established so far. Therefore, the diagonal lines connecting p-nwKRDS with Npl-Rjb and Sjp in Figure 7.20 are given without arrowheads —which indicates that the sequencing is ambiguous. There are three alternative

[28] The Maoist insurgency of the time did not permit lengthy field trips for data collection.

possibilities: (i) the Npl-Rjb and Sjp PNs may have emerged as distinct networks after the western KRDS stage, but were subsequently reintegrated into the north-west KRDS PN; or (ii) the north-west KRDS PN may have diverged directly from the western KRDS stage, and was then subsequently divided into the Surjapuri and Nepal Rajbanshi PNs; or lastly (iii) all three PNs (nwKRDS, Npl-Rjb, Sjp) may have co-existed within a complex north-west KRDS speech community during the Modern KRDS period. This historical problem is unsolved at present.

7.5.2.2 The influence of Hindi on the north-west KRDS speech community

The influence of Hindi on this subgroup of lects is diagnosed by [PI 29.], [MI 10.], [PI 19.] and [PI 34.]. RL and KS are slightly more affected by diglossic relations with Hindi than MH, as shown by [MI 8.] (cf. §7.5.1). The impact of diglossia on these lects is largely phonological, with some changes also in case marking. The verbal morphology has been unaffected by diglossia.

7.5.3 Extended central KRDS

The lects classed as 'extended central KRDS' are TH, SH, RP and BH. These lects represent the Kamta/Rajbanshi/Deshi Bhasha lects of today's West Bengal (excluding north Dinajpur) and Bangladesh. These lects have undergone PEs which define them as a Propagation Network and subgroup (§7.5.3.1). They are also characterised by diglossia with SCB as the superposed variety (§7.5.3.2).

7.5.3.1 The extended central KRDS propagation network

This PN is not very robustly supported by PEs, hence it is prefixed by '?' in Figure 7.20. The innovations relevant to diagnosing this PN are:

[PI 14.] *l > /n/ / #_ {RP, SH, BH, and TH Hindus}. Supportive, not diagnostic.

[PI 15.] *r > Ø / #_ {RP, variably in SH and BH, and among TH Hindus}. Diagnostic.

[MI 42.] *bɛ > Ø in ANP and REL temporal pronominals {TH, SH, RP, BH, BN}. Diagnostic.

The inclusion of BN in [MI 42.] is slightly problematic because otherwise BN does not show any indication of membership within the extended central KRDS PN. Rather, BN's inclusion in [MI 42.] can be explained by a distinct propagation network between BH and BN (diagnosed on other grounds, see §7.5.4.1). The dialectological ranges of [PI 14.] and [PI 15.] are given above in Figures 7.13 and 7.15 respectively. These ranges have been assigned to the Modern KRDS period (after the river shift and district organisation) in §7.4.1.2.

7.5.3.2 The influence of SCB on the extended central KRDS speech community

The lects within the extended central KRDS subgroup share a similar linguistic ecology, characterised by a diglossic relation with SCB. The innovations which result from diglossia across the extended central KRDS speech community are [PI 28.] and [MI 9.],

and three more changes occur just in TH: [MI 54.], [MI 55.] and [MI 58.]. Of the eight lects included in the reconstruction in Chapters 4–6, TH is the lect most affected by contact relations with SCB during the Modern KRDS period. The effects of diglossia are most noticable in the TH verbal morphology (cf. §6.2).

Concerning the two changes that reflect broader diglossia in the area, the possibility has been mentioned in §5.3.9 that [MI 9.] may constitute an earlier retention. On balance, however, borrowing from SCB of this ablative marker is perhaps more plausible. Concerning [PI 28.], Chatterji also supports a recent chronology and contact-related explanation for the present-day range of this innovation (1926:142).

7.5.4 Eastern KRDS

Eastern KRDS has only one representative in the reconstruction in Chapters 4–6: BN. There are many innovations which separate BN from the other KRDS lects; some are unique to BN (§7.5.4.1), but most constitute convergence with SCA norms (§7.5.4.2). The boundary between central and eastern KRDS is between Gosaigaon and Bogribari (on the central side of the boundary) and Kokrajhar (on the eastern side of the boundary). This boundary has been stable throughout the middle and modern KRDS periods. As a result, the eastern KRDS changes cannot be disambiguated into different periods in history. Thus, the chronology of changes in BN's linguistic history after proto Kamta remains unknown.

7.5.4.1 The eastern KRDS propagation network

In the reconstruction of Chapters 4–6, there is only one change that is diagnostic of a Propagation Event and centres on BN:

[PI 1.] Devoicing of the obstruent element (not the aspiration) of initial voiced aspirates {regular in BN, variable in BH}. Diagnostic.

This change is well established in BN, and sporadic in BH. It is likely that this change was innovated in BN some time ago, but is now undergoing further propagation to include BH variably. Beyond this, nothing has been reconstructed of its chronology.

There are two further changes which show evidence of propagation between BH (or possibly its nearest relative RP) and BN:

[MI 42.] *bɛ > Ø in ANP and REL temporal pronominals {TH, SH, RP, BH, BN}. Diagnostic.

[MI 57.] VERB-INF + present-perfective of *lag- 'attach' > 'present continuous' {RP, BH, BN}. Supportive, not diagnostic.

These changes reflect a recent increase of interaction between speakers of RP, BH and BN lects. Sociohistorically, this increased interaction and the resultant PEs are plausibly due to the substantial migration that has occurred from northern Rangpur to Bongaigaon during the 20th century. Many of these migrants would have been speakers of a lect basically the same as RP or BH.

7.5.4.2 The influence of SCA on the eastern KRDS speech community

Asamiya lects have undergone major changes in their phonology, such as the merger of inherited dental and postalveolar stops and the fricativisation of inherited affricates; eastern

KRDS (BN) has undergone these same changes. Furthermore, eastern KRDS has undergone major convergence with Asamiya norms in its nominal and verbal morphology (cf. §6.4.3). The relevant changes formulated in Chapters 4–6 for BN are:

[PI 11.] Apical series > alveolar articulation {BN and SCA} (during or before C15th in Asamiya, C20th in BN). Diagnostic of contact relations with SCA through diglossia.

[MI 4.] > /-[ɔ]r/ 'GEN' {BN, from SCA}. Supportive of contact relations with Asamiya.

[MI 11.] > /pɔra/ 'ABL' {BN, SCA}. Diagnostic of contact relations with Asamiya.

[MI 12.] > /kɔi/ 'CMP' {BN, SCA}. Diagnostic of contact relations with Asamiya.

[MI 25.] /mɔj, tɔj/ 'I, you' {BN}. Diagnostic of contact relations with Asamiya.

[MI 28.] > /ami/ 'we' {BN}. Supportive, not diagnostic, of contact relations with Asamiya.

[MI 30.] > /apuni/ '2H.NOM', /apona-/ '2H.OBL' {BN}. Supportive, not diagnostic, of contact relations with Asamiya.

[MI 35.] > /-ba/ 'INDF' in pronouns {BN, from SCA}. Diagnostic of contact relations with Asamiya.

[MI 72.] > /-a/ '2H' {BN} (chronology uncertain). Diagnostic of contact relations with Asamiya.

With the exception of [PI 11.], whose recent occurrence in BN has been discussed in §4.3.6, there seem to be no criteria (linguistic, textual or sociohistorical) for sequencing the rest of these changes. The linguistic boundary between Gosaigaon-Bogribari and Kokrajhar does not coincide with a clear geographical or social boundary. Historically, the border between Assam and the Koch kingdom, and then later between Assam and the Mughal empire, moved back and forth between the Bar Nadi (much further east than Kokrajhar) and the river Sankosh (further west from Gosaigaon). Neither of these historical social divisions accounts for the location of the divide between central and eastern KRDS. The question remains: why has SCA exercised such a high degree of influence over the KRDS lect of Bongaigaon and Kokrajhar, but not over the lect spoken in the adjacent areas of Dhubri, Gosaigaon, and Bogribari (all still within the State of Assam)? The sociohistorical conditioning of this consistent line of interrupted propagation between central and eastern KRDS is presently unknown, and left for further research.

7.6 Conclusion

This chapter has outlined a coherent account of the history of KRDS based on rigorous sociolinguistic reconstruction of the sequencing of linguistic changes. The sociohistorical criteria for linguistic sequencing have supplemented the linguistic and textual criteria that were applied in earlier chapters. The final results of the historical reconstruction are depicted by an adjusted tree diagram in Figure 7.20. By way of final conclusion, the next chapter surveys the methodological, social and historical linguistic implications of the study as a whole.

8 *Conclusions and implications*

This study has, from the outset, been undertaken with two kinds of reader in mind. First, the historical linguist: interested primarily in the discussion and application of theory and method. Second, the educated speaker of a Kamta/Rajbanshi/Deshi Bhasha lect: wanting to understand more about his/her mother tongue, its history, and its relation to other lects. My hope is that the study has been somewhat satisfactory for both kinds of reader, and also for those readers in whom the two interests may be combined.

In this final chapter, conclusions and implications are presented in two sections, in order to address separately the concerns of these two kinds of reader. The conclusions presented in §8.1 speak to the concerns of KRDS speakers; those in §8.2 focus on the interests of historical linguists. The intention is not to repeat the historical argument of preceding chapters but to reflect on implications of this reconstruction for the contemporary status of KRDS on the one hand, and for historical linguistic methodology on the other.

8.1 Conclusions for speakers of Kamta/Rajbanshi/Deshi Bhasha

It has not been the object of this study to construct a proof one way or the other on the controversial question of the contemporary status of KRDS. As argued in §1.3, the debate whether KRDS is a 'distinct language' or a 'dialect of Bangla' is for many on both sides not really about language at all; rather it is about social and *political* identity. Language status is used as a political symbol of social status—as a symbol of sociopolitical *autonomy* versus *subordination*—and consequently, when it comes down to it, the debate is primarily about the status of the *speakers* (autonomous or subordinated?) not the status of the *language.* The two issues are interlinked; but it remains helpful to remember the distinction. The debate on the sociopolitical status of speakers is a necessary one, but this is not the place for it. Therefore, it is my intention to remain as focussed as possible in this section on the issues surrounding the status of the language, and not enter into much discussion of the sociopolitical status of its speakers.

Nevertheless, this much must be said: history has implications for the present, and a confused understanding of the past can lead to distorted thinking about the present realities. Furthermore, influential people with sociopolitical agendas (of whatever colour) can promote accounts of language history without proper consideration of the historical veracity of their statements. The simple villager, of course, has no such recourse to promote his own opinions in reply. A careful and critical reconstruction of history is

necessary in order to protect the marginalised from the pseudo-historical ideologies which the powerful may wish to promote in order to justify their position.

The following conclusions may be justified on the basis of the historical reconstruction of preceding chapters:

(1) A stage of linguistic history, termed in this study as 'proto Kamta', is a justified historical reality. It is defined by linguistic changes that (as best can be determined) occurred between AD 1250 and 1550 in the community centred on Kamtapur—the relocated capital of the Kamrupa kingdom. This much is a *historical linguistic* statement, based on reconstruction of the chronology of linguistic changes. The chosen label of 'proto Kamta' is also a historical statement, and is not intended as a justification for any contemporary political party.

The linguistic history reconstructed here shows that all KRDS lects—whether the 'Rajbanshi' of Morang district in Nepal, the '*sthaniyo bhasha*' (local language) of Rangpur in Bangladesh, or the 'Kamta' of Cooch Behar in India—share a common ancestor, which for historical reasons may be termed proto Kamta. This common linguistic ancestor is not a fantasy created to justify a contemporary political position, but a historical entity reconstructed by the best historical linguistic methodology available to us. On one occasion when collecting data with KRDS speakers in a somewhat remote village, we were interrupted by a local official protesting that the people are 'simply making up the language that you are recording' and 'no-one speaks this way here'—in short the official claimed that the lects in question *do not exist*. After taking information regarding the purpose of my research the official left, and, a little shaken, I returned to collecting the data from speakers whose linguistic tradition is no fantasy, but as argued in this study, almost eight centuries old.

(2) This historical stage, proto Kamta, is reconstructed as historically parallel, not subordinate, to the historical emergence of proto Bangla and proto Asamiya from the common Magadhan stage. The implication of this statement is that the KRDS lects reflect a linguistic (and cultural) tradition equally as ancient as the Bangla and Asamiya linguistic traditions. This study thus confirms Clark's proposal, following Henry Frowde, that 'Northern Bengali may be as old or older than standard Bengali' (1969:85), and Grierson's statement that 'Northern Bengal and Assam did not get their language from Bengal proper, but directly from the west' (Grierson 1903–28:vol.1:126).

(3) Kamta/Rajbanshi/Deshi Bhasha *is not bad, or corrupted Bangla*. Statements to this effect by numerous 19th and early 20th century researchers (cf. those quoted in van Driem 2001) are simply a distortion of the historical reality. I find myself on the point of digression into the social status of the speakers, but will restrict myself to posing the following question: is it right that children who speak KRDS lects are ridiculed at school for using their linguistic norms, some of which are more ancient than the norms of Standard Colloquial Bangla? I intend no political statement whatsoever by this question, but merely to illustrate the fact that *the status of KRDS lects is historically misrepresented in contemporary north Bengal society*. See also the resolutions given in §1.3.

(4) While KRDS may have replaced an earlier Tibeto-Burman language, the shift from non-Aryan to Indo-Aryan language is *far from being unique to KRDS* and is no justification for ascribing inferior status to the KRDS lects. Writing with regard to Bengali, Klaiman finds that historical language shift is responsible for the bulk of that contemporary speech community: 'It is a reasonable hypothesis … that descendants of non-Bengali tribals of a few centuries past now comprise the bulk of Bengali speakers' (Klaiman 1990:512). It is inconsistent then to use historical language shift as a reason to treat the Kamta/Rajbanshi/Deshi/ Surjapuri lect(s) as inferior to Bengali.

(5) Since the 16th century (during the middle and modern KRDS periods), KRDS lects have not existed in isolation from broader Magadhan and NIA changes. During this period, central KRDS in particular has undergone changes in common with Bangla and Asamiya. The KRDS, Bangla and Asamiya communities have in general *not been closed off from one another*. That is, while their linguistic traditions are distinct from one another, they are yet intertwined, and not separated.

(6) Furthermore, since the 16th century, the KRDS lects have undergone differentiation from one another. This diversification is not merely the result of 'contamination' with other languages; it is natural in human society that lects spoken by differentiated speech communities tend to become linguistically differentiated. Many speakers tell me that 'in Nepal they speak differently to us because they mix with Maithili' or 'in Bengal they speak differently to us because they mix with Bangla', etc. etc. There is some truth to these statements, but they are far from being the whole truth of the matter. Some proto Kamta features are maintained in parts of Nepal, or Bengal, but have been changed elsewhere; some innovative and unique features have sprung up in each of these different regions, and are unrelated to the ecologies of language contact. Linguistic differentiation is just part and parcel of linguistic history. A consequence of this is that when speakers in Nepal and Cooch Behar use quite different linguistic varieties in songs, videos, and newspapers, this is not a denial of their linguistic history, but because of it. Their history has a common origin 500 years ago, but since then there has been much diversification—to the point where lects at different points in the continuum share low inherent intelligibility without acquired bilingualism. As a result, there are today *two* distinct standards emerging in the literature of KRDS speakers. The variety of *central Jhapa* features in an increasing number of publications aimed at speakers in Nepal. The variety of *eastern Cooch Behar* is likewise increasingly used for publications aimed at the Rajbanshis and deshi Muslims of northern West Bengal and western Assam.

(7) The standardisation of Bangla, Asamiya, Nepali and Hindi, and the propagation of these standardised varieties during the 19th and 20th centuries has had significant effects upon the KRDS lects. The influence of Hindi upon Nepal KRDS (Rajbanshi) and Bihar KRDS (Surjapuri) has been reconstructed in §7.5.2.2. Similarly, the influence of standard colloquial Bangla upon KRDS lects in West Bengal and northern Bangladesh has been reconstructed. The most significant influence due to language contact is that of standard colloquial Asamiya upon the Koch Rajbanshi spoken in Bongaigaon and Kokrajhar of Assam (cf. §7.5.4.2).

(8) The absence of an early standardised form of KRDS used in written literature is not simply the fault of external powers. As I have undertaken this reconstruction of linguistic history it has struck me that patronisation of Bangla and Asamiya written varieties by the Koch kings—rather than the mother tongue of their subjects—during the middle and modern KRDS periods is a major reason why these lects have been subsequently accorded the status of 'dialect' of either Bangla and Asamiya. When Grierson categorised 'Rajbanshi' as a 'dialect of Bangla', I am quite sure that this was based on (a) the Indo-Aryan character of the lect; coupled with (b) the absence of a considerable written literature in the lect; and (c) the patronisation of written Bangla and Asamiya varieties by the Koch Kings. (Unlike Chauduri 1939, Grierson does not seem to have given much importance to oral literature when categorising Indo-Aryan lects).

(9) Given this *modus operandi* behind the handing out of 'dialect' status in the *Linguistic Survey of India*, this status is not irreversible. With the development of an increasing written literature in the KRDS varieties of Jhapa and of Cooch Behar, the question of recognition becomes worthy of reconsideration. Let me reiterate that I am commenting here on the social recognition of the language, not the sociopolitical status of speakers. The situation may be illuminated if we consider that at one time the status of 'dialects of Bengali' was also accorded to Asamiya and Oriya. Mohanty (1982:22–23) describes the following episode from 1869 during what he terms 'the Bengali language dispute':

> Dr. Mitra [an eminent Bengali historiographer] asserted that the population of Orissa being barely 20 lakhs [2 million] it would be an absurdity to maintain a separate language for so few people … In the course of the agitation attempts were not only made to prove that the Oriya language did not have a separate identity but books and articles were printed which distorted the history of the land. In 'Utkal Hitaisini' (the periodical of the domiciled Bengalis) it was said that this land owed its development in religion, language and administration to Bengal.

As demonstrated by the present recognition of Asamiya and Oriya, the status of 'dialect' once given need not be irrevocably binding. The further example could be given of the recent recognition of Maithili.

(10) A further lesson which can be learnt from the history of relations between Assam and Bengal is that controversial relationships between different language communities need not always remains so. Chatterji (1963) wrote that 'at least one Bengali scholar settled in Assam … has sought to make partial atonement for the injustice done to Assam's language by serving the same language by his literary and other publications in it'. Similar demonstrations of *rapprochement* between linguistically distinct communities are to be welcomed also in the case of relations between the Bangla and Kamta/Rajbanshi/Deshi Bhasha speech communities.

8.2 Conclusions for historical linguists

The findings of the present study are now evaluated for an audience of historical linguists, with a focus on the success or otherwise of the innovative methods which played a large part in this reconstruction. As these statements will be (hopefully) less controversial

than those in §8.1, they are outlined in a more summarised form and without too much hedging of the point.

(1) While textual documentation of a language can aid the reconstruction of its linguistic history, the absence of such documentation does not negate the possibility of reconstruction. This point may seem so obvious as to go without saying to historical linguists working in, for example, the Austronesian family of languages—whose history has been quite thoroughly reconstructed despite the absence of written corpora. However, as has been mentioned several times in this study, historical studies in Indo-Aryan have almost exclusively focussed on written varieties, using texts of different eras as the fixed points in establishing the chronology of linguistic changes. Reconstruction of the histories of unwritten Indo-Aryan lects has rarely been attempted.

(2) Sociohistorical criteria for sequencing changes contribute more to historical reconstruction than linguistic criteria, because most innovations are linguistically independent of each other. That is, of all the changes that a set of languages undergoes, only a few logically require a relative chronology due to bleeding and feeding of linguistic conditions. Other criteria must be used for sequencing, and in the absence of historical documentation (and perhaps even when it is present), sociohistorical criteria are the best (and may be the only) option available to the historical linguistic.

(3) Sociohistorical sequencing, and sociolinguistic reconstruction, need not be an *ad hoc* approach but can be formalised on the basis of a sociohistorical theory of language change. It has been one of the goals of this study to develop such a procedure and demonstrate both its theoretical well-foundedness and empirical usefulness.

(4) A sociohistorical theory of language change, when applied to methodology of historical reconstruction, can substantially increase our ability to reconstruct linguistic history. This *theory* of change has been accepted by historical linguists for quite some time now, but little effort has been made to bring our *methodologies* in line with the *theory*. This study has (a) developed the sociohistorical theory so as to draw out the connections with reconstruction methodology, and (b) re-articulated the reconstruction methodology so as to reflect explicitly the sociohistorical theory.

(5) The re-articulated methodology is not *complementary* to the family tree model, but rather *subsumes* that model as well as others. This approach does not negate previous reconstruction which has assumed a family tree-like shape to linguistic history, but accounts both for why the tree diagram works in some cases, and doesn't work elsewhere. Where linguistic history is family-tree 'shaped' it is because the SCEs were characterised by division, or, because the historical linguist has chosen to only reconstruct propagation events involving the divisions of speech communities. The present approach subsumes such analyses, but also opens up new possibilities of analysis where division of lects is non-discrete; where propagation events occur through *reintegration* of speech communities, as well as through division.

(6) Sociolinguistic reconstruction depends on a robust reconstruction of linguistic innovations, and a consideration of the phylogenetic diagnostic value of each individual innovation. Therefore, *historical linguistic reconstruction must precede sociohistorical linguistic reconstruction*.

(7) Sociohistorical sequencing of changes depends upon (a) disjunctions in the ranges of changes, and (b) sociohistorical or geographical phenomena which coincide with at least some of the disjunctive ranges. To the extent which either of these factors is lacking, the method will be less successful.

(8) Reconstruction which stops at the linguistic innovations must either: (a) depend on textual evidence to establish chronology of innovations (e.g. Chatterji 1926); (b) slip in unexamined sociohistorical assumptions about the 'normalcy' of SC division in the guise of a family tree model of change (e.g. Pattanayak 1966); or (c) conclude with a dialectological map instead of a coherent account of linguistic history (e.g. Maniruzzaman 1977). By viewing linguistic history through the lens of sociohistorical propagation of changes between speakers, the strengths of each of these three approaches are given a cohesive framework within which they can be integrated.

Interaction between speakers is the mechanism by which propagation of linguistic change occurs, and reconstruction of linguistic history is understandably more successful when founded on this principle.

References

Acharyya, N.N., 1966, *The history of medieval Assam (from the thirteenth to the seventeenth century)*. Gauhati: Omsons publications.

Ahmed, Khan Chowdhuri Amantulla, 1936, *Cooch Beharer Ityhas (a history of Cooch Behar, in Bengali)*. Cooch Behar: Published under the authority of the Cooch Behar State.

Aikhenvald, Alexandra Y. and Robert M.W. Dixon, eds, 2001, *Areal diffusion and genetic inheritance: problems in comparative linguistics*. Oxford: Oxford University Press.

Andrews, Avery, 1985, The major functions of the noun phrase. In Timothy Shopen, ed. *Language typology and syntactic description*, 62–154. Cambridge/New York: Cambridge University Press.

Anttila, Raimo, 1972, *An introduction to historical and comparative linguistics*. New York: Macmillan.

—— 1989, *Historical and comparative linguistics*. Amsterdam/Philadelphia: John Benjamins.

Bandyopadhyay, Asitkumar, 1991, আঞ্চলিক বাংলা ভাষার অভিধান *anchalik bangla bhashar abhidhan (a dictionary of dialectal Bengali language)*, vol. I. Kolkata: Calcutta University Press.

Barma, Dharma Narayan, 1991, *A step to Kamata-Behari language*. Adhikari: Minati Barma.

—— 2000, কামতাপুরী ভাষা-সাহিত্যের রুপরেখা *Kamtapuri bhasa-sahitter ruparekha*. Tufanganj: Raydak.

Barua, B.K., 2003, *A cultural history of Assam: early period* (4th edn). Guwahati, Assam: Bina Library.

Baruah, Sanjib and Colin P. Masica, 2001, Assamese. In Jane Garry and Carl R. Galvez Rubino, eds *Facts about the world's languages: an encyclopedia of the world's major languages, past and present*, 43–47. New York: H.W. Wilson.

Bhakat, Dwijendra Nath, 2004, Origin, formation and development of Rajbansi language. In Dwijendra Nath Bhakat, ed. রাজবংশী ভাষা প্রসঙ্গ *rajbanshi bhasha prasanga*, 37–47. Golakganj, Assam: Centre for Ethnic Studies and Research (CESR).

Bhattacharya, Pramod Chandra, 1975, Sino-Tibetan (Boro) elements in Assamese, Bengali and north-eastern Indic languages. *Indian Linguistics* 36/3:240–246.

Bhattacharya, Sudhindra Nath, 1943, State of Bengal under Jahangir. In Ramesh Chandra Majumdar and Jadu Nath Sarkar, eds *The history of Bengal*. Dacca: University of Dacca.

Bhattacharya, Tanmoy, 2001, Bangla. In Jane Garry and Carl R. Galvez Rubino, eds *Facts about the world's languages: an encyclopedia of the world's major languages, past and present*, 65–71. New York: H.W. Wilson.

Bhattacharyya, Sudhindra Nath, 1929, *A history of Mughal North-east Frontier Policy*. Calcutta: Reprinted 1994 by R.N. Bhattacharya, Calcutta, and 1998 by Spectrum, Gauhati.

Blench, Roger and Matthew Spriggs, 1998, General introduction. In Roger Blench and Matthew Spriggs, eds *Archaeology and language II: archaeological data and linguistic hypotheses*. One world archaeology, No.29, 209–261. London/New York: Routledge.

Bloch, Jules, 1920, *La formation de la langue marathe*. Bibliothèque de l'École des hautes études; 215, Sciences historiques et philologiques. Paris: E. Champion.

Bloomfield, Leonard, 1935, *Language*. London: George Allen and Unwin.

Boehm, Ed, 2004, Case marking in Dangaura Tharu. Unpublished MS, Kathmandu, 11 p.

Bubenik, Vit, 2003, Prākrits and Apabhramśa. In George Cardona and Dhanesh Jain, eds *The Indo-Aryan languages*, 204–249. London/New York: Routledge.

Burling, Robbins, 1999, On 'Kamarupan'. *Linguistics of the Tibeto-Burman area*, 22/2:169–171.

Cardona, George and Dhanesh Jain, eds, 2003, *The Indo-Aryan languages*. London/New York: Routledge.

Chambers, J.K. and Peter Trudgill, 1998, *Dialectology* (2nd edn). Cambridge textbooks in linguistics. Cambridge/New York: Cambridge University Press.

Chatterji, Suniti Kumar, 1926, *The origin and development of the Bengali language*. Reprinted 1985. New Delhi: Rupa.

—— 1963, *The place of Assam in the history and civilisation of India*. Calcutta: Bengal publishers, 14 Bankimm Chatterji Street.

Chaudhuri, Sambhu Chandra, 1939, Notes on the Rangpur dialect. *Indian Linguistics*, 7:297–315.

—— 1940, North Bengali Dialects: Rajshahi. *Indian Linguistics*, 7:418–431.

Choudhury, Pratap Chandra, 1966, *The history of civilisation of the people of Assam to the twelfth century A.D.* (2nd edn). Gauhati: Dept of Historical and Antiquarian Studies in Assam.

Clark, Robert Henry, 1969, A study of the religious customs and practices of the Rajbangshis of North Bengal. PhD dissertation, The Hartford Seminary Foundation, 598p.

Comrie, Bernard, 1978, Ergativity. In Winfred P. Lehmann, ed. *Syntactic Typology*, 329–394. Austin: University of Texas Press.

Croft, William, 2000, *Explaining language change: an evolutionary approach*. Longman linguistics library. Harlow, England/New York: Longman.

Damant, G.H., 1873, On the dialect of the Palis. *The Indian Antiquary*, V/II:101–102.

Das, Dhirendra Nath, 1990, The dialects of Goalpara and Kamrup: a comparative analysis. PhD dissertation, University of Gauhati, 340p.

Das, Nirmal, 2001, উত্তরবঙ্গের ভাষাপ্রসঙ্গ *Uttarbanger bhasha prasang: a collection of investigative papers on some sociolinguistic situations of North Bengal*. Kolkata: Sahitya Bihar.

Dasgupta, Dipankar, 1978, *Linguistic studies in Juang, Kharia Thar, Lodha, Mal-Pahariya, Ghatoali, Pahariya*. Calcutta: Anthropological Survey of India, Government of India.

Dasgupta, Probal, 2003, Bangla. In George Cardona and Dhanesh Jain, eds *The Indo-Aryan languages*, 351–390. London/New York: Routledge.

Dash, G.N., 1982, *Descriptive morphology of Oriya*. Santiniketan, West Bengal: Visva-Bharati Research Publications.

Datta, B, 1971, Dialects of Goalpara district. *The Assam Academy Review*, 1:53–74.

Driem, George van, 2001, *Languages of the Himalayas: an ethnolinguistic handbook of the greater Himalayan region*. Leiden/Boston: Brill.

Enfield, N.J., 2003, *Linguistic epidemiology: semantics and grammar of language contact in mainland Southeast Asia*. London/New York: Routledge.

Eppele, John and Jonathan Grimes, 2001, The Rajbanshi of Nepal: a sociolinguistic study. Nepal National Languages Preservation Institute.

Ferguson, C.A. and M. Chowdhury, 1960, The phonemes of Bengali. *Language*, 36:22–59.

Ferguson, Charles A., 2000, Diglossia. In Li Wei, ed. *The bilingualism reader*, 65–80. London: Routledge.

Gait, Edward Albert, 1905, *A history of Assam*. Repr. Calcutta 1963: Thacker Spink and Co.

Geraghty, Paul A., 1983, *The history of the Fijian languages*. Honolulu: University of Hawaii Press.

Gordon, Raymond G., Jr., ed., 2005, *Ethnologue: Languages of the World* (15th edn). Dallas, Texas: SIL International. Online version: http://www.ethnologue.com/.

Goswami, Golok Chandra, 1966, *An introduction to Assamese phonology*. Poona: Deccan College Postgraduate and Research Institute.

Goswami, Golok Chandra and Jyotiprakash Tamuli, 2003, Asamiya. In George Cardona and Dhanesh Jain, eds *The Indo-Aryan languages*, 391–443. London/New York: Routledge.

Goswami, Upendranath, 1970, *A study on Kamrupi: a dialect of Assamese*. Gauhati: Department of Historical and Antiquarian Studies.

—— 1974, The language of north Bengal and western Assam. *Journal of the University of Gauhati*, vol. 1.

Grierson, George Abraham, 1877, Notes on the Rangpur dialect. *Journal of the Royal Asiatic Society*, XLVI/I:185–226.

—— 1887, *Seven grammars of the dialects and subdialects of the Bihari language, spoken in the province of Bihar, in the eastern portion of the North-western Provinces, and in*

the northern portion of the Central Provinces. Part VIII. Repr. Varanasi 1980: Bharatiya.

—— 1903–28, *Linguistic survey of India*. Calcutta, Repr. Delhi 1967: Motilal Banarsidass.

Habib, Irfan, 1982, *An atlas of the Mughal Empire: political and economic maps with detailed notes, bibliography and index*. Delhi: Oxford University Press.

Haldar, Gopal, 1986, *A comparative grammar of East Bengali dialects*. Calcutta: Puthipatra.

Hodgson, Brian Houghton, 1880, *Miscellaneous essays relating to Indian subjects*. London: Trubner.

Hull, David L., 1988, *Science as a process: an evolutionary account of the social and conceptual development of science*. Science and its conceptual foundations. Chicago: University of Chicago Press.

Huson, Daniel H. and David Bryant, 2006, Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution*, 23/2:254–267.

Hutchins, Edwin and Brian Hazlehurst, 1995, How to invent a shared lexicon: the emergence of shared form-meaning mappings in interaction. In Esther Goody, ed. *Social intelligence and interaction*, 53-67. Cambridge: Cambridge University Press.

Islam, Safiqul, 1992, রাজশাহীর উপভাষা *Rajshahi upobhasha (the dialect of Rajshahi)*. Dhaka: Bangla Academy.

Jha, Subhadra, 1958, *The formation of the Maithili language*. Repr. New Delhi 1985: Munshiram Manoharlal.

Joshy, Rose Chand and Maina Joshy, 2006, A write-up on Kamta/Rajbongshi: Phonemic summary. Unpublished MS, Cooch Behar.

—— 2007, Kamta/Rajbongshi grammar. Unpublished MS, Cooch Behar, 57 p.

Kakati, Banikanta, 1962, *Assamese, its formation and development* (2nd edn). Revised and edited by Golockchandra Goswami. Reprinted 1995. Gauhati: Lawyer's Book Stall.

Katre, Sumitra Mangesh, 1968, *Problems of reconstruction in Indo-Aryan*. Simla: Indian Institute of Advanced Study.

Keller, Rudi, 1994, *On language change: the invisible hand in language*. Trans. Brigitte Nerlich. London/New York: Routledge.

Khondakar, Abdur Rahim, 1998, বরেন্দ্র ভুমির ভাষা (The language of the land of Barendra). In Md. Momtazur Rahman, A.K.M. Zakaria, Saifuddin Chowdhury et al., eds বরেন্দ্র অঞ্চলের ইতিহাস *Barendra anchaler itihas* (*a history of the Varendra region)*. Rajshahi: Office of the Divisional Commissioner.

Kirch, Patrick Vinton and Roger Curtis Green, 2001, *Hawaiki, ancestral Polynesia: an essay in historical anthropology*. Cambridge, UK/New York: Cambridge University Press.

Klaiman, M.H., 1990, Bengali. In Bernard Comrie, ed. *The world's major languages*, 490–513. New York/Oxford: Oxford University Press.

Koch, Harold, 1996, Reconstruction in Morphology. In Mark Durie and Malcolm D. Ross, eds *The comparative method reviewed: regularity and irregularity in language change*, 218–263. New York: Oxford University Press.

Koul, Omkar N., 2003, Kashmiri. In George Cardona and Dhanesh Jain, eds *The Indo-Aryan languages*, 895–952. London/New York: Routledge.

Labov, William, 1965, On the mechanism of linguistic change. In Charles W. Kreidler, ed. *Georgetown University monograph series on languages and linguistics, No.18*. Reprinted in John J. Gumperz and Dell Hymes, eds *Directions in sociolinguistics: the ethnography of communication.* (2nd edn, 1986), 512–538. London: Basil Blackwell.

—— 1966, *The social stratification of English in New York City*. Washington: Center for Applied Linguistics.

—— 1994, *Principles of linguistic change: internal factors*. Oxford: Blackwell.

—— 2001, *Principles of linguistic change: social factors*. Oxford: Blackwell.

Ladefoged, Peter, 1971, *Preliminaries to linguistic phonetics* (rev. edn). Chicago: University of Chicago Press.

Ladefoged, Peter and Ian Maddieson, 1996, *The sounds of the world's languages*. Phonological theory. Oxford, UK/Cambridge, MA: Blackwell.

Lass, Roger, 1997, *Historical linguistics and language change*. Cambridge studies in Linguistics, No.81. Cambridge: Cambridge University Press.

LeDoux, Joseph E., 1998, *The emotional brain: the mysterious underpinnings of emotional life*. New York: Simon and Schuster.

Lehmann, Winfred P., 1992, *Historical linguistics: an introduction* (3rd edn). London/New York: Routledge.

LePage, Robert B. and Andrée Tabouret-Keller, 1985, *Acts of identity: creole-based approaches to language and ethnicity*. Cambridge/New York: Cambridge University Press.

Linguistic Society of America, 1997, Resolution on the Oakland 'Ebonics' Issue. Unanimously adopted at the annual meeting of the LSA, Chicago, January 3. Available online: http://www.lsadc.org/info/lsa-res-ebonics.cfm

Lynch, John, Malcolm Ross and Terry Crowley, 2002, *The Oceanic languages*. Curzon language family series; 1. Richmond, Surrey: Curzon.

Majumdar, Durgadas, 1977, Koch Bihar, West Bengal District Gazetteers.

Majumdar, Paresh Chandra, 1970, *A historical phonology of Oriya*. Calcutta: Sanskrit college.

Majumdar, Ram Prasad, 1955, A dark chapter in the history of Bengal. *Journal of the Bihar and Orissa Research Society*, XLI/Part I:59–67.

Majumdar, Ramesh Chandra and Jadu Nath Sarkar, 1943, *The history of Bengal*. Dacca: University of Dacca.

Maniruzzaman, 1977, Controlled historical reconstruction based on five Bengali dialects. PhD dissertation, University of Mysore, 410p.

Marshall, Jonathan, 2004, *Language change and sociolinguistics: rethinking social networks*. Palgrave Macmillan.

Masica, Colin P., 1991, *The Indo-Aryan languages*. Cambridge language surveys. Cambridge/New York: Cambridge University Press.

—— 1992, South Asia as a sociolinguistic area? In Edward C. Dimock, Jr., Braj B. Kachru and Bh. Krishnamurti, eds *Dimensions of sociolinguistics in South Asia: Papers in memory of Gerald B. Kelley*. New Delhi: Oxford and IBH.

Matisoff, James A., 1999, In defense of Kamarupan. *Linguistics of the Tibeto-Burman Area*, 22/2:173–182.

Mendoza-Denton, Norma, 2002, Language and identity. In J.K. Chambers, Peter Trudgill and Natalie Schilling-Estes, eds *The handbook of language variation and change*. Malden, MA/Oxford: Blackwell.

Milroy, James, 1992, *Linguistic variation and change*. Language in Society, 19. Oxford: Blackwell.

—— 1997, Internal vs external motivations for linguistic change. *Multilingua*, 16/4:311–323.

Milroy, Lesley, 1987, *Language and social networks* (2nd edn). Oxford: Blackwell.

Misra, Haripriya, 1975, *Historical Oriya morphology*. Varanasi: Bharata Manisha.

Mohanty, Nivedita, 1982, *Oriya nationalism: quest for a united Orissa 1866–1936*. New Delhi: Manohar.

Morrison, David A., 2005, Networks in phylogenetic analysis: new tools for population biology. *International Journal for Parasitology*, 35:567–582.

Mufwene, Salikoko S., 2001, *The ecology of language evolution*. Cambridge Approaches to Language Contact. Cambridge: Cambridge University Press.

Nathan, Mirza, II, 1936, *Baharistan-i-Ghaybi: a history of the Mughal Wars in Assam, Cooch Behar, Bengal, Bihar and Orissa during the reigns of Jahargir and Shahjahan*. Trans. Moayyidul Islam Borah. Shillong: Government of Assam, Department of Historical and Antiquarian Studies.

Ngwazah, Liahey, Scott Turner and Matthew Toulmin, 2006, A Sociolinguistic survey of the Kamta, Rajbanshi, and Surjapuri languages and dialects in West Bengal, Assam and Bihar states of India. New Life Computer Institute.

Nichols, Johanna, 1996, The comparative method as heuristic. In Mark Durie and Malcolm D. Ross, eds *The comparative method reviewed: regularity and irregularity in language change*, 39–71. New York: Oxford University Press.

Pal, A.K., 1966, Phonemes of a Dacca dialect and the importance of tone. *Journal of the Asiatic Society of Bengal* 7:1–2.

Pattanayak, Debi Prasanna, 1966, *A controlled historical reconstruction of Oriya, Assamese, Bengali and Hindi*. Janua linguarum: Series practica, No.31. The Hague: Mouton & Co.

Pawley, Andrew and Malcolm D. Ross, 1993, Austronesian historical linguistics and culture history. *Annual Review of Anthropology* 22:425–459.

—— 1995, The prehistory of Oceanic languages: a current view. In Peter Bellwood, James J. Fox and Darrell Tryon, eds *The Austronesians: historical and comparative perspectives*, 39–74. Canberra: Dept. of Anthropology, The Australian National University.

Pawley, Andrew and Timoci Sayaba, 1971, Fijian dialect divisions; eastern and western. *Journal of Polynesian Society* 80/4:405–436.

Pischel, R., 1981 [1900], *A grammar of the Prakrit languages*. Trans. Subhadra Jha. Delhi: Motilal Banarsidass.

Poudel, Tika Ram, 2006, *Rajbanshi grammar and interlinearized text*. LINCOM Studies in Indo-European Linguistics, No.34. München: Lincom Europa.

Purkait, Nikhilesh, 1989, বাংলা অসমীয়া ও উড়িয়ার উপভাষার ভৈগোলিক জরিপ (*Dialect geography of Bengali, Assamese, and Oriya)*. Calcutta: Subarnarekha.

Ramachandran, Subhadra, 2001, Oriya. In Jane Garry and Carl R. Galvez Rubino, eds *Facts about the world's languages: an encyclopedia of the world's major languages, past and present*, 528–530. New York: H.W. Wilson.

Ray, Tapas S., 2003, Oriya. In George Cardona and Dhanesh Jain, eds *The Indo-Aryan languages*. London/New York: Routledge.

Rennell, James and Brahmadeva Prasad Ambashthya, 1975, *Memoir of a map of Hindustan or the Mughal Empire and his Bengal atlas*. Patna: N.V. Publications.

Ross, Malcolm D., 1988, *Proto Oceanic and the Austronesian languages of western Melanesia*. Canberra: Pacific Linguistic.

—— 1997, Social network and kinds of speech-community event. In Roger Blench and Matthew Spriggs, eds *Archaeology and language I: theoretical and methodological orientations*. One world archaeology, No.27, 209–261. London/New York: Routledge.

—— 1998, Sequencing and dating linguistic events in Oceania: the linguistics/archaeology interface. In Roger Blench and Matthew Spriggs, eds *Archaeology and language II: archaeological data and linguistic hypotheses*. One world archaeology, No.29, 141–173. London/New York: Routledge.

Ross, Malcolm D. and Mark Durie, 1996, Introduction. In Mark Durie and Malcolm D. Ross, eds *The comparative method reviewed: regularity and irregularity in language change*, 3–38. New York: Oxford University Press.

Sanyal, Charu Chandra, 1965, *The Rajbansis of North Bengal: a study of a Hindu social group*. Asiatic Society, Calcutta. (Monograph series vol. 11). Calcutta: The Asiatic Society (Monograph series vol. 11).

Sarkar, Jadu Nath, 1943, Transformation of Bengal under Mughal rule. In Ramesh Chandra Majumdar and Jadu Nath Sarkar, eds *The history of Bengal*. Dacca: University of Dacca.

Saussure, Ferdinand de, 1966, *Cours de linguistique générale* (3rd edn). Paris: Payot.

Sen, Sukumar, 1960, *History of Bengali literature*. New Delhi: Sahitya Akademi.

—— 1971, *An etymological dictionary of Bengali, c. 1000-1800 A.D*. Calcutta: Eastern.

Shahidullah, Muhammad, 1966, *Buddhist mystic songs: oldest Bengali and other eastern vernaculars* (rev. edn). Dacca: Bengali Academy.

Shapiro, Michael C., 2003, Hindi. In George Cardona and Dhanesh Jain, eds *The Indo-Aryan languages*, 250–285. London/New York: Routledge.

Shapiro, Michael C. and Harold F. Schiffman, 1981, *Language and society in South Asia*. Series in linguistics, No.1. Delhi: Motilal Banarsidass.

Shukla, Shaligram, 2001, Bhojpuri. In Jane Garry and Carl R. Galvez Rubino, eds *Facts about the world's languages: an encyclopedia of the world's major languages, past and present*. New York: H.W. Wilson.

Sircar, Dines Chandra, 1952, Spread of Aryanism in Bengal. *Journal of the Asiatic Society of Bengal, Letters,* XVIII/2:171–178.

Southworth, Franklin C., 1958, A test of the comparative method (a historically controlled reconstruction based on four modern Indic languages, Panjabi, Hindi, Bengali and Marathi). PhD dissertation, Yale University.

—— 1964, Family-tree diagrams. *Language* 40/4:557–565.

Spate, O.H.K. and B.H. Farmer, 1954, *India and Pakistan: a general and regional geography* (1st edn). London: Methuen.

Sperber, Dan, 1996, Anthropology and psychology: towards an epidemiology of representations. *Man* 20:73–89.

Tagare, Gajanan Vasudev, 1948, *Historical grammar of Apabhramsa*. Deccan College dissertation series, No.5. Poona, India: Deccan College.

Thomason, Sarah Grey and Terrence Kaufman, 1988, *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Thurgood, Graham, 1999, *From ancient Cham to modern dialects: two thousand years of language contact and change*. Oceanic linguistics special publication; No.28. Honolulu: University of Hawai'i Press.

Tiwari, Udai Narain, 1960, *The origin and development of Bhojpuri*. Kolkata: Asiatic society.

Toulmin, Matthew, 2002a, A grammar sketch of the Rajbanshi language as spoken in Mahayespur, Jhapa, Nepal. Unpublished MS, Department of Linguistics, University of Sydney, 65p.

—— 2002b, Mui Tui Bangla, Rajbanshi or Kamatapuri? A study of language, ideology and classification in north Bengal. Honours dissertation, Department of Linguistics, University of Sydney, 114p.

—— 2006, Reconstructing linguistic history in a dialect continuum: the Kamta, Rajbanshi, and Northern Deshi Bangla subgroup of Indo-Aryan. PhD dissertation, School of Language Studies, The Australian National University. Available online: http://thesis.anu.edu.au/public/adt-ANU20070411.000201/index.html

—— in press, Rajbanshi/Kamta, as spoken in India and Bangladesh. In Tatiana Oranskaia, ed. *New Indo-Aryan languages*. Languages of the World. Moscow: Nauka.

Trask, R.L., 2000, *The dictionary of historical and comparative linguistics*. Edinburgh: Edinburgh University Press.

Trudgill, Peter, 1974, Linguistic change and diffusion: description and explanation in sociolinguistic dialect geography. *Language in Society* 3:215–246.

—— 1986, *Dialects in contact*. Language in society, No.10. Oxford/New York: Basil Blackwell.

Turner, Ralph Lilley, Sir, 1962–66, *A comparative dictionary of Indo-Aryan languages*. London: Oxford University Press. Accessed online at http://dsal.uchicago.edu/dictionaries/soas/.

Verma, Manindra K., 2003, Bhojpuri. In George Cardona and Dhanesh Jain, eds *The Indo-Aryan languages*. London/New York: Routledge.

Verma, Manindra K. and Karuvannur Puthanveettil Mohanan, 1990, *Experiencer subjects in South Asian languages*. Stanford, CA: Published for the Stanford Linguistics Association by the Center for the Study of Language and Information Stanford University.

Verma, Sheela, 2003, Magahi. In George Cardona and Dhanesh Jain, eds *The Indo-Aryan languages*. London/New York: Routledge.

Voegelin, Charles Frederick and Florence Marie Robinett, 1977, *Classification and index of the world's languages*. Foundations of linguistics series. New York: Elsevier.

Weinrich, Uriel, William Labov and Marvin I. Herzog, 1968, Empirical foundations for a theory of language change. In W.P. Lehmann and Y. Malkiel, eds *Directions for historical linguistics*, 95–195. Austin: University of Texas Press.

Whyte, Brendan R, 2002, *Waiting for the Eskimo: an historical and documentary study of the Cooch Behar enclaves of India and Bangladesh*. Melbourne: The School of Anthropology, Geography and Environmental Studies, University of Melbourne.

Wilde, Christopher Pekka, 2002, Preliminary phonological analysis of Rajbanshi with notes on orthography problems. Unpublished MS, Kathmandu, 34p.

—— 2008, A sketch of the phonology and grammar of Rājbanshi. PhD dissertation. Publications of the Department of General Linguistics, University of Helsinki, 43.

Wolfram, Walt and Natalie Schilling-Estes, 2003, Dialectology and linguistic diffusion. In Joseph and Janda, eds *Handbook of Historical Linguistics*, 713–735.

Yadav, Ramawatar, 2003, Maithili. In George Cardona and Dhanesh Jain, eds *The Indo-Aryan languages*. London/New York: Routledge.

Zograph, G.A., 1976, *Morfologicheskii stroi novykh indoariiskikh iazykov*. Moscow: Nauka.