A History of the Bhojpuri (or "Hindi") Language in South Africa

by

Rajend Mesthrie

Thesis submitted for the degree of Doctor of Philosophy
in the Department of Linguistics
at the University of Cape Town
1985

Supervisor: Prof. R. Lass

The University of Cape Town has been given the right to reproduce this thesis in whole or in part. Copyright is held by the author.

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

CONTENTS

ABSTRACT

Although Indian languages have existed in South Africa for the last 125 years, there are no academic studies of any of them - of their use in South Africa, their evolution and current decline. Many misconceptions persist concerning their names, their structure, and status as 'proper' languages.

This thesis deals with the history of one such language, Bhojpuri (more usually, but incorrectly, referred to as "Hindi"). I attempt to trace the origins of the South African variety of this language by examining the places of origin of the original indentured migrants who brought it to South Africa. A complex sociolinguistic picture emerges, since these immigrants came from a very wide area in North India spanning several languages. I also attempt to describe the early history of Bhojpuri in South Africa as a 'plantation' language. Subsequent changing patterns of usage are then detailed, including phonetic, syntactic, lexical and semantic change. The influence of other South African languages - chiefly English, but also Zulu, Fanagalo, and other Indian languages - is described in detail, as well as changes not directly attributable to language contact. A final section focusses on the decline of the language and the process of language death.

From another (more international) perspective this study lays the foundation for comparisons between Bhojpuri in South Africa and other 'overseas' varieties of it, spawned under very similar conditions, in ex-colonies like Surinam, Fiji, Mauritius, Guyana, Trinidad and others. Such a comparative study could well make as great a contribution to general and socio- linguistics as the study of creoles has in the recent past.

Information concerning this unwritten language was gathered by field-work throughout Natal. This involved informal interviews with over two hundred

fluent speakers, including four who had been born in India during the time of immigrations. The study also draws upon the author's observations on language practices as an 'inside' member of the community under study.

ACKNOWLEDGEMENTS

I am truly grateful to all those who served as informants for this study, for responding to my intrusions and queries with patience and interest, for sharing with me the story of their lives, and for the warm hospitality which always included at least a warm cup of tea. I regret that it is not possible to name each one individually (for there are too many of them).

Among the many friends and relatives who offerred me sustenance and accommodation during field-trips all over Natal, I must thank Ramesh and Premilla Mungal of Pietermaritzburg, Mr and Mrs Badal of Stanger, Mr J. Pandit, also of Stanger, Mr and Mrs B. Maharaj of Ladysmith, Mr and Mrs H. Durgadin, Mr and Mrs B.C. Maharaj, and Mrs G.R. Maharaj, all of Newcastle, and the Mistrey families in Dannhauser.

Thanks are due to Mrs Bhaw and Mrs H.M. Singh of Durban for assistance in interviewing the more elderly informants, and to my former students who assisted in gathering valuable data: 'Bubbles' Juggan, Reena Singh, and Shakila Maharaj. To this list I must add Radhika Chowthee for data from remoter parts of Natal.

In Patna, Bihar I profited from discussions with Profs K.M. Tiwari and B. Varma, who also assisted me organise some field-trips. I also thank their families for their hospitality. Prof R.N. Srivastava of Delhi University also gave me the benefit of his knowledge of Indian Socio-linguistics. Communication with scholars working on Bhojpuri in other parts of the world have proved stimulating - I thank Surendra Gambhir, Peggy Mohan, Sita Kishna, and Kunti Ramdutt for their interest in this study.

A work like this invariably depends on others for its very existence: my reliance on Gambhir's work (in chap 3), on Dorian (in chap 6), on Tiwari (in chap 2), and Sir George Grierson (throughout) will be apparent. In addition I have found Peggy Mohan's thesis on Trinidad Bhojpuri valuable, as well as Domingue's study of Mauritian Bhoj, which first brought Bhojpuri to my attention. Haugen's study of the Norwegian

language in America was an early source of inspiration.

I gratefully acknowledge a doctoral grant from the Human Sciences Research Council, which assisted in covering part of my expenses, and an ad hoc grant from the Anglo-American Chairman's fund on very short notice, which enabled me to attend a conference at the University of the West Indies. The views expressed in this study are not necessarily those of either body.

The staff at the Don Africana and Killie Campbell libraries in Durban deserve special mention for their efficienct services. I also thank Prof A. Brimer of the English department at Durban-Westville for organising my teaching schedule in such a way as to provide me with time for field-work. I have been lucky enough to have had many good teachers at school and university, and would like to thank three great scholars for giving me more encouragement than they realise: Profs P. Pillay (Dept of Mathematics - UDW), W.P. Lehmann (Dept of Linguistics - Austin, Texas) and J.M. Coetzee (Dept of English - UCT).

The typing was the corporate effort of many talented persons: Zarina Khan, Yogie Joseph, and Jaithri Singh, while the less professional bits are all mine. I am very grateful to Priya Banwari for the meticulous care with which he re-drew my once ungainly maps.

My family put up good-humouredly with my eccentricities during my studies: thanks are due to Priscilla, Vidia, Vijay, Shobha, and especially my mother, who checked out my linguistic intuitions and helped me interpret some of my data. My wife, Uma who endured my absences from home during field-work, and long periods of silence during the writing of this thesis, was a constant source of support. I am grateful to her, and to my mother, and mother-in-law, Mrs Dhupelia for keeping me well-fed, even though I did little to deserve it.

Many friends helped in small but invaluable ways to make life bearable by relieving me of administrative and other chores on many occasions: I particularly thank Vishnu, Damian, Connie, Shireen, Nadia, and Kirti.

This study has benefitted from comments by Susan Wright and Richard Bailey. I am particularly grateful to Dr Sitaram of the Hindi department at Durban-Westville for graciously checking my data and analyses, and saving me from error on several occasions. Most of all, I thank my supervisor, Prof. R. Lass of the department of Linguistics at UCT for the extremely prompt and close attention given to my work, for sharing with me his expertise in Linguistics and his considerable editorial skills, and for his encouragement from start to finish. All errors and shortcomings in this work are, however, to be laid at my own door.

SYMBOLS

In this text IPA symbols are used for phonetic descriptions, but a slightly different system of transliteration, based on western descriptions of Sanskrit and modern Indic languages, is used otherwise. The following is a list of the symbols and letters used in transliteration, their IPA and Devanagari equivalents, and a brief description of their values. (Symbols like (b), (e), and (s) whose IPA and transliterated values coincide are excluded).

Transliteration	IPA equivalent	Devanāgarī symbol	Description	Example from SABh
t	[t̪]	त	voiceless dental stop	tār 'wire'
d	[d̪]	द	voiced dental stop	dānā 'grain'
ś	[ʃ]	श	voiceless palatal fric.	śēṭ 'shirt'
c	[tʃ]	च	voiceless alveopalatal affric.	cār 'four'
ch	[tʃʰ]	छ	voiceless alveopalatal aspirated affric.	chū- 'to touch'
j	[dʒ]	ज	voiced alveopalatal affric.	juttā 'shoe'
jh	[dʒʰ]	झ	voiced alveopalatal aspirated affric.	jhār- 'to sweep'
f	[f̬]	फ़	voiceless labiodental approximant	safā 'clean'
v	[ʋ]	व	voiced labiodental approximant	Ved (a proper name)
y	[j]	य	palatal semi-vowel	maiyā 'mother'
h	[h]	ह	voiceless glottal approx. (at end of syllables)	patōh 'daughter-in-law'
	[ɦ]		murmured glottal approx. (elsewhere)	hanḍi 'pot'
ph	[pʰ]	फ	voiceless bilabial aspirated stop	phūl 'flower'
kh (etc)	[kʰ] (etc)	ख	voiceless velar aspirated stop (etc)	khūn 'blood'
bh	[bɦ]	भ	murmured bilabial stop	bhāi 'brother'
gh (etc)	[gɦ] (etc)	घ	murmured velar stop (etc)	ghāū 'a sore'
ṭ	[ʈ]	ट	retroflex voiceless stop	ṭangi 'axe'
ḍ	[ɖ]	ड	retroflex voiced stop	ḍar 'fear'

Transliteration	IPA equivalent	Devanagari symbol	Description	Example from SABh
ṇ	[ɳ]	ण	retroflex voiced nasal	thaṇḍa 'cold'
m̃	[ɲ]	ञ	palatal nasal	pañc 'public'
ṛ (etc)	[ɽ] (etc)	ड़	retroflex voiced flap (etc)	baṛa 'big'
ī	[i:]	ई	long high front vowel	tīn 'three'
ō (etc)	[o:] (etc)	ओ	long mid back vowel (etc)	khōl- 'open'
ā	[ɑ:]	आ	long low back vowel	khā- 'eat'
ã	[ɑ̃]	अँ	nasal low back vowel	hãs- 'laugh'
m̃̄ (etc)	[ɑ̃:] (etc)	आँ	long nasal low back vowel (etc)	ghm̃̄as 'grass'
â or á	[ɵ]	-	mid central rounded vowel	(Indian Bhoj dekhá- 'see'

The following symbols also occur in the text:

[ɬ]	Voiceless alveolar lateral fric.	Zulu hlakula 'to hoe'
[ɫ]	velarised lateral	English call
[θ]	voiceless dental fricative	English thigh
[ð]	voiced dental fricative	English thy
[ʒ]	voiceless alveopalatal affric.	English pleasure
[ɾ]	voiced alveolar flap	RP very
[ɹ]	voiced alveolar approximant	RP red
[ə]	unstressed mid central vowel	English ago
[æ]	low front vowel	RP cat
[ɔ:]	rounded mid back vowel	RP hall
[ɒ]	rounded low back vowel	RP hot
[ɨ̈]	centralised /i/	SAE bit
[ɐ̈]	murmured /ɑ/ (etc)	SABh bhai 'brother'
[i̥]	voiceless /i/ (etc)	Indian Bhoj ãkhi 'eye'

ABBREVIATIONS

ABL	Ablative
ACC	Accusative
AD	Adessive
Adj	Adjective
affric	affricate
approx	approximant
asp	aspirate
aux	auxiliary
Awad	Awadhi
Bhoj	**Bhojpuri**
BPL	Bihar Peasant Life
CAUS	Causative
Cent	Central
CF	Counter-factual
CLASS	Classifier
COM	Comitative
Comp	Complement
Conj	Conjunctive
CORR	Correlative
DAT	Dative
DEF	Definite
DET	Determiner
E	Eastern
Eng	English
fem	feminine
fut	future
fric	fricative
GEN	Genitive
H	High
HAB	Habitual
Hn	Hindi
IMP	Imperative
INCL	Inclusive
INF	Infinitive
INTERROG	Interrogative
intrans	intransitive
IPA	International Phonetic Alphabet
L	Low
L1	first language
lit	literally
LOC	Locative
LSI	Linguistic Survey of India
Maga	Magahi
Maith	Maithili
masc	masculine
MIA	Middle Indic
n	noun
No	number
NOM	Nominative
NP	Noun Phrase
O	Object
OBL	Oblique
OFS	Older Fluent Speaker
OIA	Old Indic
Past P	Past Participle
pl	plural
Pn	Pronoun
PP	Present Participle
pres	present
+R	honorific
-R	non-honorific
REFLEX	Reflexive
REL	Relative
RP	Received Pronunciation
S	Subject
SA	South Africa
SABh	South African Bhojpuri
SAE	South African English
SAIE	South African Indian English
sg	singular
Skt	Sanskrit
SS	Semi-speaker
Std Eng	Standard English
Std Hn	Standard Hindi
SUBJ	Subjunctive
trans	transitive
V	Verb
VN	Verbal Noun
VP	Verb Phrase
W	Western
YFS	Young Fluent Speaker

MAPS

TABLES

INTRODUCTION

Bhojpuri in South Africa is (like Tamil and Telugu) a dying language: parents have for some time ceased to ensure that the language is transmitted to their children, most of whom are now growing up as English monolinguals, with some passive knowledge of Bhojpuri. As there is a great need to record as much as possible about the life-cycle of Bhojpuri in South Africa before it becomes entirely extinct, and as I have had no previous descriptions of any Indian language in South Africa to draw upon, I have accordingly spread my nets widely in presenting an overview of the language, rather than concentrating on specific aspects of its grammar or sociolinguistics.

I grew up half-believing (like most people) that what was spoken in my home was "Hindi" in a very "broken" form, an attitude that is all too readily expressed by priests and educated people captivated by Standard Hindi, and by youngsters understandably eager for any explanation of their lack of competence in it. This study arose partly on account of my own ignorance of, and curiosity regarding, this ancestral language. Far from being unworthy of formal study, Bhojpuri presents - as I try to show in chapter 5 - a partial social history of its speakers, and a link between Indian communities as far off as Surinam, Mauritius, Fiji, the West Indies, and South Africa. Bhojpuri has been the language of indentureship par excellence, though this not to deny Tamil and the other Indian languages their rights to similar claims. It is as 'grammatical' as any other language of India or South Africa, though, as one might expect, it has evolved a few rules of its own.

This text is based primarily on field-work undertaken in Natal from January 1982 to September 1984. During this period I interviewed 182 fluent speakers from most parts of Natal, wherever there were sizeable Bhojpuri-speaking communities. The interviews were as informal as possible, and spontaneous insofar as very few were conducted by prior arrangement. (One easily identifies a Bhojpuri-speaking household by the red flag raised in the prayer section of their yard.) In each town a minimum of 15 houses were visited (and many more in the bigger cities),

including some from the newer suburbs as well as the poorest sections. In addition I was often directed by residents to 'feeder' rural areas, where older, more settled communities exist.

The interviews were usually conducted in (Indian) English, though for older informants I used Bhojpuri as well, and sometimes enlisted the aid of younger relatives of such interviewees in setting up an informal interview. Older informants (over 50) were encouraged to talk about their family histories, present and past occupations, and about any other topic that took their fancy. This proved quite successful in eliciting natural speech, with very few interviewees seeming to be inhibited by the presence of a tape-recorder.

Informants under 50, usually fluent in English, were asked to translate orally a word-list, and a series of sentences graded in terms of length (and couched in Indian English) into Bhojpuri (or "Hindi" as it is known locally). This proved useful in eliciting morphological data (especially for dialect study) and phonological information in a short space of time. No inferences were made concerning syntax from these, since some speakers showed the influence of the English phraseology in their translations. All observations of a syntactic nature in this study are drawn from spontaneous conversation or from non-translated narratives (which were obtained from as many speakers as possible, time permitting).

In another ten instances students and friends of mine made tape recordings of conversations in their homes, when there were mainly elders about, ensuring that Bhojpuri, not English, would be predominantly used. Since this was often done covertly, it provided more informal data than the interviews I had conducted - and useful information on code-switching.

These figures exclude 15 younger (and less fluent) speakers, for whom a slightly different interview procedure and questionnaire were used (described in chap 6). Finally by noting down, privately, patterns of usage in the homes of relatives and other acquaintances I was able to gather much information regarding all aspects of Bhojpuri grammar, and particularly semantic change and the use of loanwords. In this way I was

able to circumvent the observer's paradox (that one tries to observe how people behave when they are not being observed).

The size and structure of the sample drawn upon in chapter 3 (concerning dialectal forms in lexis and morphology) is as follows:

a) Total number interviewed: 182

b) Place of Residence:

Upland Natal	25%
Midland Natal	37%
Coastal Natal	37%
Rural Natal	26%
Urban Natal	74%

c) Sex:

Female	70%
Male	30%

d) Age:

Over 70	7%
60 - 69	19%
50 - 59	34%
40 - 49	19%
30 - 39	17%
under 30	4%
(No information)	1%

e) Occupation:

Housewives	61%
Blue Collar Jobs	15%
Clerks	5%
Private Businesses	5%
Farming	4%
Priests	4%
Teachers	3%
Unemployed	3%

f) Hindi education: (part-time)

4 years or more	16%
1 - 3 years	17%
No education	65%
(No information)	2%

Certain imbalances in the sample were unavoidable. Women (especially housewives) were more easily available for interviews than men. Men in outside employment by day often proved reluctant interviewees. As Bhojpuri is more often spoken in the home, and by women (see chapter 4, where a detailed analysis of speech networks within twenty families is presented), the male-female ratio in the sample might not be an unrealistic one.

Although the figures indicate many more urban interviewees than rural ones, one should take into account that many of the former had been brought up in a rural environment. The low percentage of informants under the age of thirty reflects the rarity of fluent speakers in this age-group - see chapter 6. The level of education in English of the informants was not fully analysed; the majority, however, had only four to six years of full time schooling, while many older interviewees had had no formal education. Only a few of the younger ones had had a high school or college education.

CHAPTER 1

THE HISTORICAL BACKGROUND

1.1 The Migration to Natal: "Hindi", as indicated by official sources, is one of several Indian languages existing in South Africa since 1860, the time of the first immigration. The name "Hindi" is, as my quotation marks suggest, problematic, but that is an issue we must defer till later in this chapter. The advent of Indians under the indentured immigration scheme is amply recorded (see, for example, Thompson 1952, Palmer 1957, and Kuper 1960). A brief summary follows:

Organisation of a system of indentured immigration of Indians began in the 1830's, after the abolition of slavery in the British Empire (1834) had brought about a shortage of manual labour in many colonies. The Indian government, entirely British-administered at that time, permitted the recruiting of Indian labourers by the sugar-planting colonies under certain conditions. Five-year contracts between the labourers and the colonial powers had to be drawn up; a Protector of Emigrants had to be appointed in each colony to supervise the terms of indenture; and recruitment of labourers had to be performed by natives of India, under the control of agents appointed by the colonies. Emigration was permitted from three ports only - Calcutta, Bombay, and Madras. Colonies more distant than Mauritius had to guarantee a free return passage to workers who had served there for at least ten years (Thompson 1952:11). Under these conditions, thousands of Indians were shipped, first to Mauritius (1842), then to Jamaica, British Guyana, and Trinidad (1844), and later to St. Lucia (1856), and Grenada (1858). In 1860 there were shipments to Reunion, Martinique, Guadeloupe and French Guyana, (all French Colonies), St Croix and Surinam, (both Dutch colonies), and Natal (a British colony). Recruitment to Fiji and Australia followed later.

In 1860 sugar-cane planters in Natal faced a shortage of labour, as

recruitment of local Zulus, more or less self-sufficient in their tribal areas, had not proved successful. On the recommendation of the Governor of the Cape Colony, Sir George Grey, who had seen the flourishing cane-fields of Mauritius, Natal planters agreed to the introduction of indentured Indian labour. The first ship, Truro, carrying migrant labourers from Madras and its environs arrived on November 16th, with 203 men and 140 women and children on board. They were followed a few days later by the Belvedere from Calcutta, with 351 passengers. The departure of these, and subsequent emigrants, was not always voluntary. Rev. C.F. Andrews, an associate of Mahatma Gandhi, and himself very much involved in the welfare of the early Indians in Natal, said this of the indentured immigration system:

> ... it was promoted and controlled by Government .... Professional recruiters, who were paid a high price for each recruit, were licensed by the Government to go in and out among the village people in order to induce them to leave their homes and be sent abroad for the purpose of labour. This kind of immigration was all too frequently accompanied by deception on a large scale ..... It cannot be stated too clearly that such immigration is artificial in the extreme. It must never be mistaken for the natural flow of the Indian People to foreign lands. Had it not been for the eagerness of the British Colonies to obtain cheap labour for their plantations, it would never have taken place at all. Indians would have stayed at home. It was Natal, Mauritius, Fiji, etc which wanted the government of India to send Indian labour, and not vice-versa. (in Sannyasi and Chaturvedi 1931:28)

At the same time, it must be conceded that for many of the migrants, employment in Natal held some flickering hope of escape from debt, failing crops, or unemployment.

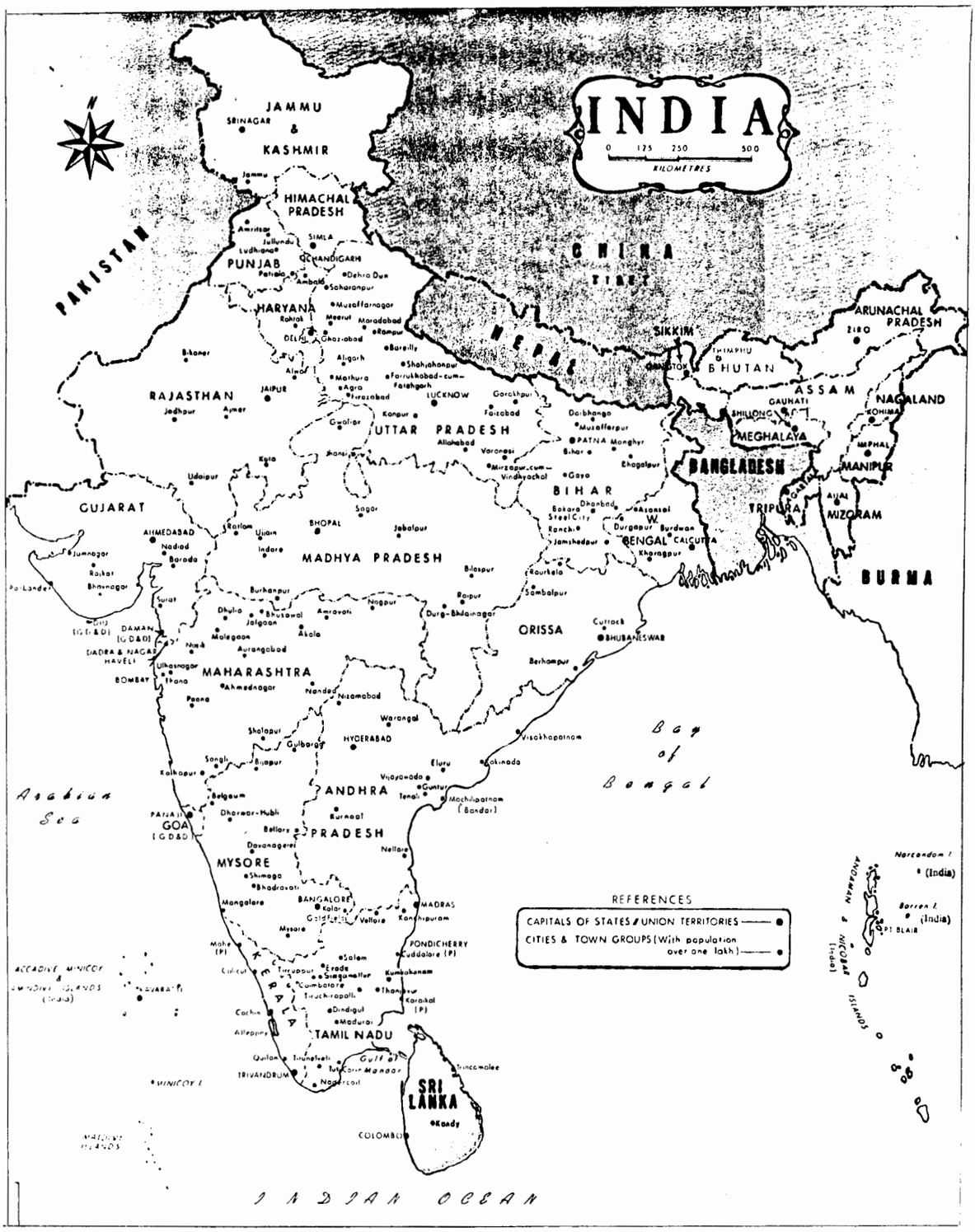
The period of migration to Natal stretched from 1860 to 1911, with a temporary halt between 1866 and 1874. During this period 152,184 indentured workers had made the journey to a new life, the peak period of migration being between 1874 and 1885. The following table gives some relevant statistics for the period 1860 to 1886:

YEAR	CALCUTTA	MADRAS	TOTAL
1860	312	601	913
61	240	359	599
62	0	0	0
63	0	668	668
64	384	1 857	2 241
65	0	984	984
66	0	864	864
67-73	0	0	0
74	4 310	0	4 310
75	2 057	0	2 057
76	754	0	754
77	1 089	1 173	2 262
78	1 722	3 576	5 298
79	0	1 116	1 116
80	505	1 168	1 673
81	1 703	909	2 612
82	872	753	1 625
83	1 457	947	2 404
84	1 337	1 626	2 963
85	389	850	1 239
86	0	227	227
	17 131	17 678	34 809

Table 1 - Number of indentured Indian workers entering South Africa, 1860 to 1886. (Source:Meer 1980:311-312).

The 1911 Census records report 149,791 Indians resident in South Africa in that year - 93,886 of whom were male. 63,776 of the total Indian population had been born on South African soil.

These bonded labourers soon found out that the new land was not the worker's paradise that many recruiters had made it out to be. Conditions of work in the early years were most unsatisfactory, as the high suicide rate bears out. Long working hours, inequitable male-female recruitment patterns, flogging, overcrowding in barracks, and poor pay



Map 1 - Present-day India
(source: India - A Reference Annual 1973).

have led more than one historian to remark that the only difference between the conditions of slavery and indentureship was that the latter was not for life (see eg. Tayal 1980:30-n; Palmer 1957:44)

Initially this class of indentured labourers occupied the lowest rungs of Natal's socio-economic ladder - the word coolie contemptuously denoting, at once, 'Indian' and 'indentured worker'. After the initial period of indenture was over (five years, later extended to ten), these labourers were theoretically free to seek employment elsewhere, to re-indenture, to return to India (which a small proportion did), or to take up farming on small plots of land allotted to those who had served for ten years. This latter offer was withdrawn by 1891 (Kuper 1960: 2). Imposition of taxes on "free" Indians forced many to re-indenture. Unlike many other British colonies, Natal did offer opportunities for independent economic advancement: immigrants found small-scale vegetable and fruit production more profitable than labouring in the cane-fields. As early as 1866 many Indians were being employed on the railways in various menial positions. By the end of the nineteenth century there were 1317 Indian workers in the coal mines of Northern Natal, and thousands employed in jobs that were more rewarding than that of field-labourer - as waiters, hawkers, fishermen, domestics etc (Bhana and Brain 1984: part 2, 28).

Another group of Indian migrants came via Bombay to Natal, from 1875 onwards, comprising Hindu and Muslim merchants, who arrived voluntarily with the intention of setting up small businesses in various parts of the province. These were the so-called 'passenger Indians', whose relative wealth and trading interests kept them apart socially, and sometimes politically, from the indentured Indian.

The subsequent history of Indians in South Africa records the enormous influence of Mahatma Gandhi in their early politics, the fight for a less degrading status and greater economic rewards, the struggle against the threat of repatriation (only in 1961 was permanent citizenship unambiguously guaranteed to South Africa's Indians), and against second-class citizenship and disenfranchisement in a segregated society.

Today very few Indians are still to be found in the canelands, and the descendants of the original settlers hold diverse positions, ranging from manual labourers to white-collar workers and large-scale entrepreneurs (some of whom themselves own small sugar estates). The financial outlook has improved over the last few decades, together with the level of education; the scale ranging from very poor illiterates, mainly in country districts, to the very affluent in a few "select" suburbs. Whereas most 'passenger' Indians (traders), still have close ties with India, both economic and social, to the extent of encouraging marriage between South African-born youngsters and natives of India, until this was banned by South African law in 1956, the descendants of indentured labourers have virtually no contact with the villages their forefathers left behind. A visit to India is for most of them a fairly recent experience, made possible by an improved economic position. Those who have recently holidayed in India express mixed emotions about the experience - typically mixing awe for the grandeur of certain aspects of Indian religious life and culture, with feelings of alienation from crowded cities and rural poverty.

The many families who chose to return to India in the thirties and forties did not find it easy to re-assimilate to village life with its fine caste gradations and other restrictions. The original workers who migrated to South Africa had not all been of the same caste. There were members of all four caste-groupings: Brahmin (priests), Kshatriya (Warriors), Vaishya (Merchants), Sudra (workers), though the majority, not surprisingly, tended to be of relatively "low" caste. Kuper (1960: 7) estimates the latter to be sixty percent.[1] The new economic situation that the indentured labourer and his family found themselves in necessitated great changes in the caste-structure.

Whereas caste rules in India determined whom one socialised with, married, ate with, where one ate, what one ate, to whom one offerred food, how one prayed, dressed and spoke, in Natal people of all castes were expected to labour together, live in close proximity to each other, and share common facilities. Upon serving his contract the ex-

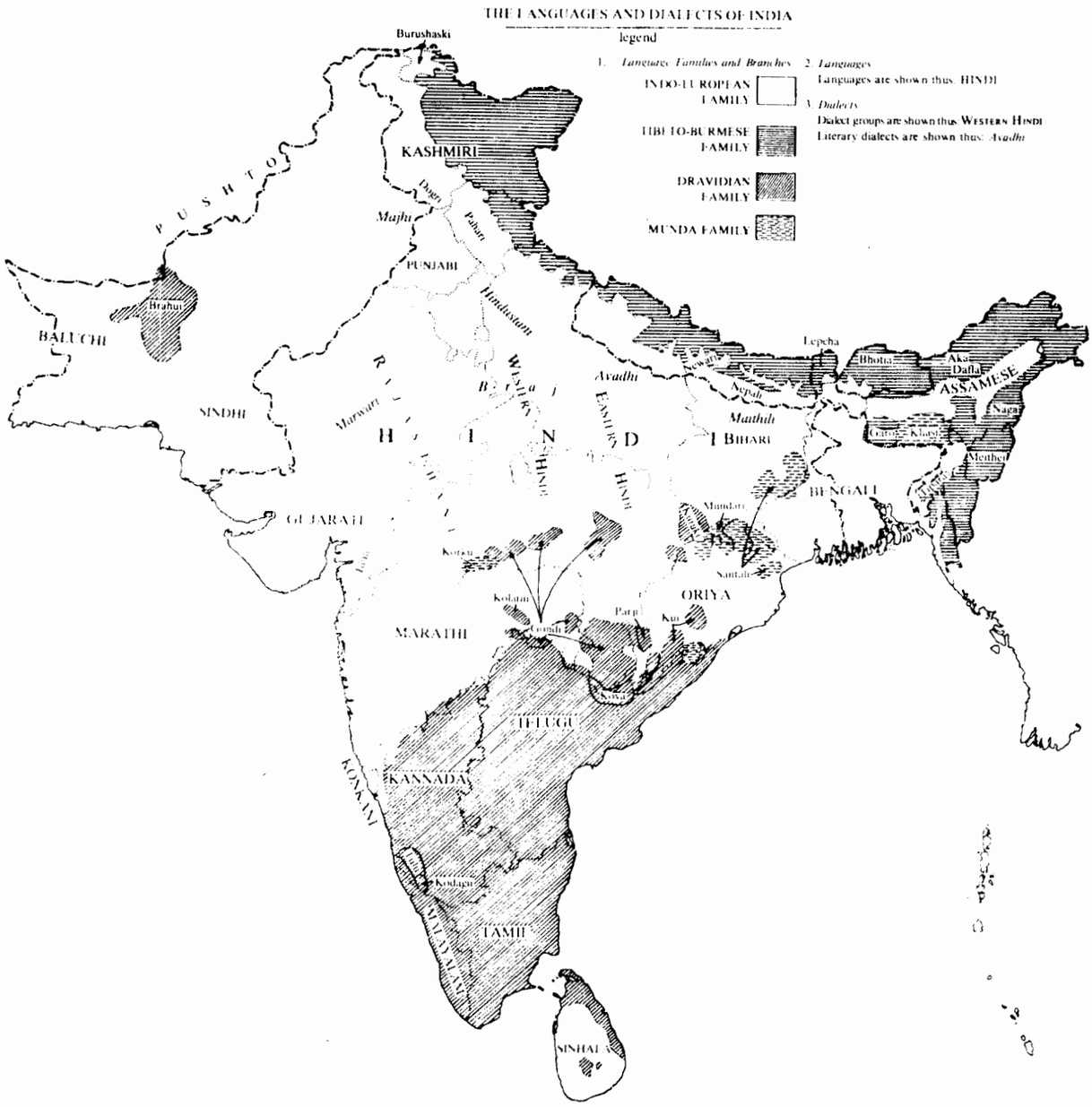
indentured labourer could not simply return to his traditional caste-occupation, because that occupation presupposed a well-defined village system in which the role of one inhabitant complemented that of the others, in a fine network of co-operation. In addition, some caste-occupations were rendered obsolete in the new land - the oil-maker, popcorn-maker, and shepherd all had to look for more viable means of livelihood. Today caste lingers on as a memory, and there has been a levelling-out of caste distinctions for all practical purposes (in terms of work opportunities, place of residence etc.), though it does manifest itself in the occasional ban on marriage by parents, between a so-called (and vaguely-defined), "high caste" and "low-caste" couple (usually by the "high-caste" family).

A broad classification of the language families of India follows, as a prelude to the discussion of the linguistic affiliations of the South Africa-bound migrants.

1.2 The languages of India: In present-day India there are, according to the 1961 Census, 723 Indian languages, divided into 1652 dialects. There are also 103 non-Indian languages, including important ones like English and Arabic. However only 50 of the Indian languages have over a hundred thousand speakers, 14 of which have more than ten million speakers each[2]. These languages (excluding immigrant languages of the recent past) can be divided into four major families:

a) The Indic family, which covers most of north and central India, and includes Hindi-Urdu, Bengali, Panjabi, Kashmiri, Gujarati, Marathi etc., all of which are traceable eventually to Sanskrit - that Indo-European language brought to India around 1500 B.C., or to related dialects - the Prakrits.

b) The Dravidian family, which is today restricted mainly to the south, comprising Tamil, Telugu, Kannada, and Malayalam (all written languages) and many unwritten languages such as Toda, Kota, Tulu etc., some of



Map 2 - The Languages of India
(source: Kachru 1983)

which are in danger of extinction today. These languages all descend from a common ancestor language dubbed Proto-Dravidian.

c) The Munda family, which is perhaps the oldest language family in India, and includes languages like Santali, Mundari, and Ho - spoken by so-called "tribals", and Nicobarese, spoken on the Andaman islands. It is thought to be related to the Austro-Asiatic languages outside India - to Khmer, Malay etc.

d) The Tibeto-Burman family, which exists in the Himalayan region and areas bordering on Burma, comprising languages like Bhote, Khasi, and Ladhaki. In terms of the number of languages it contains, this is the most prolific family in India, though in terms of the total number of its speakers it is only third, almost fourth. It is generally accepted as belonging to a larger Sino-Tibetan family outside India.

Though historically unrelated and quite dissimilar in structure and vocabulary three thousand years ago, these languages have over time converged on account of their mutual influence over each other. As a simple example we can cite the word for 'chilli' or 'pepper' being roughly the same in most Indian languages today - originating from the Munda maridsa, thence into Sanskrit (marica), and Dravidian (eg Modern Tamil milaku). The word for 'curry' seems to have originated with Dravidian (Modern Tamil kari), and thence spreading throughout, later even into English. It is therefore possible to speak of most of India today as a single linguistic area[3], following Emeneau (1956). Some of the major similarities are:

i) Phonetic: Most languages of India have a set of retroflex consonants, pronounced with the tip of the tongue curled back to strike hard against the back of the alveolar ridge, in contrast to a dental series, involving the tongue against the upper teeth. This retroflex series is, most certainly, an example of the Dravidian family's influence over the sub-continent.

In addition most Indian languages have a voiceless series of aspirates, differentiating them from their unaspirated counterparts. This is a

characteristically Sanskritic feature which eventually spread to other language families, though colloquial Tamil and Malayalam remain notable exceptions.

ii) Non-Phonetic features: All the languages of India are verb-final languages, with (S)OV order. This is a particularly important structural phenomenon, since from it follows a host of other syntactic properties: in verb-final languages one expects adjectives to precede nouns, auxiliaries to follow verbs, one expects postpositionns, not prepositions, etc.

Characteristic of most Indian languages is an echo-construction, in which a word is repeated with the initial syllable only changed, to convey a sense of "collective-ness". Thus the Tamil phrase puli-gili means 'tigers and suchlike', derived from the word puli 'tiger', with gili being the echo. An example from Hindi would be ghar-or 'houses etc' where or is the echo of ghar 'house'. This construction seems to be either Munda or Dravidian in origin (Emeneau 1956).

On the other hand both these families have in the more recent past assimilated certain structural features from the Indic languages, including possibly the use of some complex sentence patterns and simplified types of negation (see Andronov 1964).

1.2.1 The Indic languages: As our concern in this study is with some languages belonging to the Indic family, a brief outline of the history of this group of languages might be useful. Linguists outline three stages in the history of Indic (also known as Indo-Aryan): Old Indic, Middle Indic, and New Indic (often abbreviated to OIA, MIA, and NIA).[4]

Old Indic: This is the term for the aggregate of dialects spoken by those Indo-Europeans who came to India in several groups and over many generations around 1500 BC, which appear in literary form in ancient religious texts, the Vedas. Although the Vedas show signs of later editing by priests, they retain many archaic features uncharacteristic of later texts. The term Vedic Sanskrit (or just Vedic) is used to differentiate this literary language from its ensuing Sanskritic phase.

Sanskrit represents a literary development of some Old Indic dialects, and came to be the standard norm for writing, especially after Panini's grammar, the Aṣṭādhyāyī, of the fourth century BC, completely specified its rules of usage. Literature in Sanskrit stretches from about the beginning of the first millenium BC (Epic Sanskrit of the Rāmāṇaya and Mahābhārata) till well into the Middle Indic phase, Classical Sanskrit being a major written medium up to about 1000 AD.

Middle Indic: While Sanskrit remained perfected and unchanging as a literary medium for centuries, the spoken varieties of Old Indic continued changing, until by the middle of the first millenium BC we can speak of a new stage of the language - Middle Indic. Early in this period arose the Prakrits, developments of the spoken Old Indic dialects, with phonetic changes and variation in grammar that marked them off as being clearly different from them. These Prakrits were favoured by reform movements like Buddhism and Jainism because they were closer to the living speech of the majority than was Sanskrit, the language of orthodox Brahminism. Prakrit is divided into three branches: a) Śaurasenī (or western Prakrit) which was spoken in the central area between the Ganges and the Jumna, b) Māgadhī (or Eastern Prakrit) spoken in Magadha (today's South Bihar) and c) Ardha-Māgadhī (or half-Māgadhī) which is geographically and linguistically intermediate between the first two.

The Prakrits were still synthetic languages, like Sanskrit, but they show assimilation of conjunct consonants, and a great reduction in the number of case forms, noun classes, and verb forms. Towards the latter half of the Middle Indic period (i.e. circa 500 AD) the Prakrits themselves had become primarily associated with literature, and in ordinary speech the Apabhraṁśas took shape. (Etymologically Apabhraṁśa means 'corrupted or decayed speech', as opposed to Samskṛta 'perfected' or 'polished', and Prakṛta 'natural' or 'ordinary'.) They represent a spoken development of the Prakrits, but eventually came to be used for literary purposes as well, though to a lesser extent than Sanskrit. These Apabhraṁśas represent a transitional stage between the Prakrits and the modern languages of India.

New Indic: After 1000 AD differences between the various Apabhraṁśas intensified, resulting in the rise of many new vernaculars. These, too, came to be used as media for poetry and some prose: around 1200 Prithi Raj Rasan was writing in a form of Western Hindi; Narsimha Mehta's works, in an early form of Gujarati, are dated between 1415 - 1418; while Tulasi Das's version of the Rāmāyaṇa in Early Awadhi dates to the sixteenth century. The New Indic phase is characterised by languages which are more analytic than their predecessors; which have fewer case forms, use a great number of postpositions, develop a system of compound verbs, have an ergative or ergative-like construction in the past, and contain a large core of non-Sanskritic vocabulary, chiefly from Persian (which in many instances itself borrowed them from Arabic) after the Muslim conquest of India, and indigenous languages of Dravidian and Munda origin (though these are to be found in Sanskrit as well, to a lesser extent).

1.2.2 Classification of the modern Indic languages: Grierson's (1927) categorisation of the present day vernaculars into three groups, based on historical origins as well as synchronic differences is as follows:

A. OUTER SUB-BRANCH:

I. Northwestern Group

1. Lahnda (or Western Panjabi)
2. Sindhi

II. Southern Group

3. Marathi

III. Eastern Group

4. Oriya
5. Bihari
6. Bengali
7. Assamese

B. MEDIATE SUB-BRANCH:

IV. Mediate Group

8. Eastern Hindi (i.e. Awadhi, Bagheli, and Chatisgarhi).

C. INNER SUB-BRANCH:

V. Central Group

9. Western Hindi (i.e. Hindi, Dakhini Hindi, Braj, Bangaru, Kanauji, and Bundeli)
10. Panjabi
11. Gujarati
12. Bhili
13. Khandesi
14. Rajasthani

VI. Pahari Group

15. Eastern Pahari (or Nepali)
16. Central Pahari
17. Western Pahari

Table 2 - Classification of the Indic languages (after Grierson 1927).

The Inner languages descend from Śaurasenī Prakrit, and the Apabhraṁśas that arose from it, and differ from the Outer languages in having a less synthetic morphology, in their treatment of /s/, which remains as [s], whereas it is changed to [ɦ] or [ʃ] in the Outer languages, in the absence of /l/ in the past participle suffix (with the exception of Gujarati), in the use of a passive construction with the past form of transitive verbs, and others (see Grierson 1927).

The Outer languages descend from various sources: the Eastern group from Magadhi Prakrit, Marathi from Mahārāṣṭrī Prakrit (which was a sub-division of Ardha-Māgadhī Prakrit, leaning more towards Māgadhī than Śaurasenī), while Sindhi and Lahnda, whose early histories are not entirely clear, seem to be derived from Apabhramsas which show Śaurasenī influence.

The Mediate group, comprising various dialects of Eastern Hindi, is descended from Ardha-Māgadhī, sharing some features like the absence of /l/ in the past participle suffix with the Inner languages, and others like the treatment of transitive verbs via a passive construction in the past, with the Outer branch.

Grierson's classification, enormously influential, but not without its problems tied, as it is, to a Stammbaum model of language evolution, and inconsistent in its criteria for the distinctions language/dialect and what he occasionally terms "corrupt" or "mixed speech" as against "true" dialects, has been partially modified by a few scholars. Zograph's adaptation of Chatterji's (1926) refinement of Grierson is worthy of note: he discards the notion of an Inner and Outer core of languages, and repositions problematic cases like Gujarati and Bihari as set out in table 3.

1.3 The Linguistic Situation in Natal: Migrants to Natal spoke languages belonging to either the Dravidian or Indic families; there are no traces of Munda or Tibeto-Burman languages ever being used locally. From the

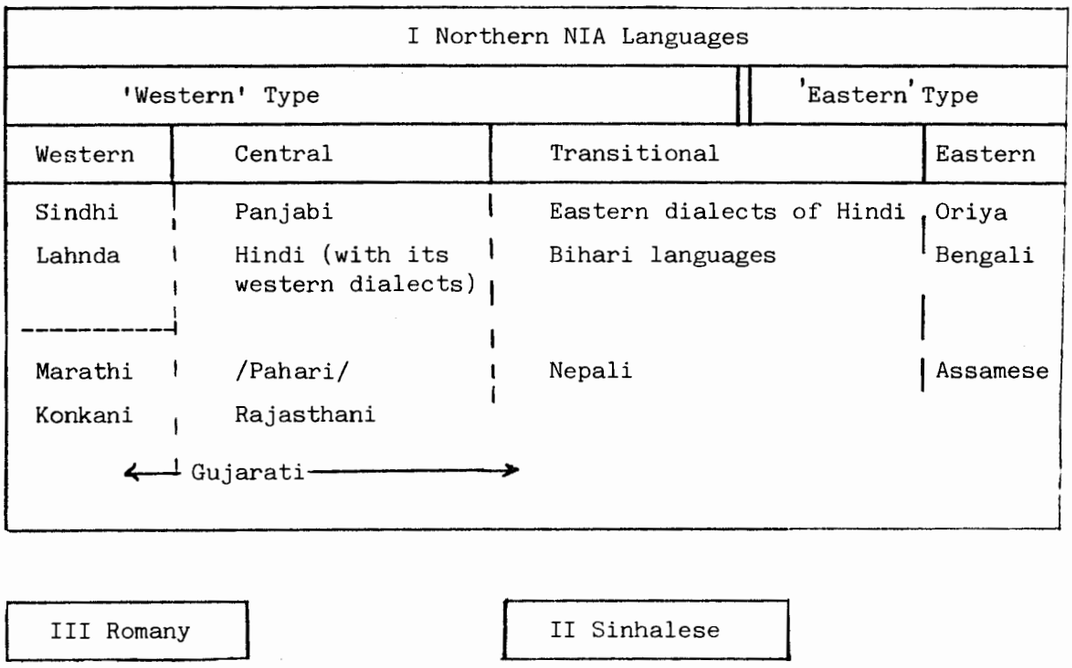


Table 3 - Classification of the Indic languages (after Zograph 1982).

south, via Madras, came the Dravidian languages Tamil, Telugu, and Malayalam, the last having only a few score speakers. From the north, via Calcutta, came "Hindi", Bengali, Panjabi, and a few others. From the west, via Bombay, came Gujarati (in particular the Surti and Kathiawadhi dialects) and Meman, a dialect of Sindhi, influenced by Gujarati. In addition there was Urdu, the language mainly of indentured Muslims, and Konkani, a language historically related to Marathi. The earliest census reports of 1911 are, unfortunately, quite unreliable for estimating the numbers of speakers of these languages [5]. By 1936 the only languages with a sizable number of speakers were, in descending order: Tamil, "Hindi", Gujarati, Telugu, and Urdu. Some figures for Indian languages, taken from Population Census records are set out below.

	1936	1951	1960	1970
Tamil	83 731	120 181	141 977	153 645
Hindi	60 276	86 594	126 067	116 485
Gujarati	25 408	39 495	53 910	46 039
Telugu	25 077	30 210	34 483	30 690
Urdu	13 842	25 455	35 789	- -
Other	2 737	26 090	2 053	71 070

Table 4 - No of Speakers of Indian languages in South Africa (1936 to 1970).

At the outset, these various groups (South Indian, North-East Indian, Gujarati, Urdu-speaking Muslim and others), were quite distinct culturally and linguistically. The Hindi-speakers of the north knew no Tamil and vice-versa. The Gujarati-speaking trader was quite often educated in both Hindi and Gujarati. English, the official language of the new land, was unknown to the majority. Several historians (see, for example Thompson 1952:47) record translation difficulties in both court and cane-field since very few interpreters had been brought over from India, and even fewer were competent enough in all of, say, English, Tamil, and Hindi.

Contact between indentured families on the fields and in the barracks resulted in some bilingualism amongst their children, especially in Hindi and Tamil. In addition, Fanagalo, a pidginised from of Zulu, with a lexicon that drew from English, Afrikaans, and Zulu - was learned by the whole Indian community, for communication with Zulus and occasionally with employers. The acquisition of English was slower, and less perfect. Women were slower than men in taking to English, lacking contact with native speakers of that language. Today in isolated communities one still finds one or two Indians (usually female), conversant in Hindi, Tamil, and Fanagalo, but not English. There are also a few cases where the only common language between two elderly

Indians, usually female, is not Tamil, Hindi, or English (say), but Fanagalo. (Detailed discussion of the early sociolinguistic situation is postponed till chapter 4.)

Until the last ten years, the various Indian linguistic groups have tended to keep apart in their social lives, as far as it was possible. In particular, marriages between persons belonging to different language groups were rare, and generally met with disapproval from the respective communities. Despite this cultural/linguistic isolation, the outlook on life of most Indians growing up in South Africa in the twentieth century has been quite uniform. One uniting factor was that education acted as a leveller, overriding linguistic and cultural differences to some extent, since it was almost entirely in English and oriented towards the West.

In the first fifty years Indian education in South Africa was characterised by its lack of system. In the early years English missionaries ran a few schools which admitted Indian pupils; subsequent progress in education was instigated by the Indians themselves, with some teachers being recruited from Mauritius and India. The medium of instruction in these schools was English, with no vernacular languages featuring at all. Education tended to be not always effective because of the high drop-out rate at an early age in favour of full-time jobs, the lack of well-educated teachers and the often tenuous relationship between subject matter emphasized in syllabi and the life style of the pupils. Reading, writing and arithmetic predominated, with a little grammar, geography and history in the larger schools.

Vernacular education was in the beginning largely oral, with traditional wisdom and knowledge being passed on by elders. Religious and epic poetry was often learnt off by heart and vernacular plays (especially in Tamil) were staged frequently. Vernacular classes were eventually introduced by various religious groups concerned with fostering the values and literature of a particular linguistic group, but met with several difficulties pertaining to staffing, the establishment of suitable classrooms and attendance. The first Tamil school, for example, was established in Durban in 1899 (Kuppusami 1946: 70). Such private schools

were never numerous at any time. Kichlu (1928) records 50 private (vernacular) schools in Natal, maintained by the Indian community, a figure which includes both full-time as well as part-time schools. Many of these (about 40, according to Kichlu) were attached to mosques, using Gujarati as medium of instruction, and in some cases Urdu. The full-time sectarian schools were eventually closed down on the recommendation of Indian educationists, and vernacular classes have since the 1930s been offered on a part-time basis in those areas where numbers warranted it.

The increase in the use of English at school was accompanied by a decline in the use of the Indian languages, even in the home. There were only a few linguistically homogeneous neighbourhoods (for example, Indians residing in the inner city of Durban are still mainly Gujarati-speaking). A Tamil-speaking household could be flanked by a "Hindi"-speaking neighbour on one side and a Telugu-speaking family on the other in most other residential areas.

The census records quoted above merely hint at the declining use of the vernaculars today. Bughwan's survey of 1970 arrives at a more realistic assessment of the situation - a clear case of a language shift currently under way by a migrant community, under pressure from a dominant national language. There have been a few efforts to stem the tide: Urdu had been taught for several years at primary level in a few schools, while Arabic is still offerred at a few high schools, for its religious value, rather than as colloquial language of any local community. After a test period of a couple of years, the vernacular languages have finally been introduced in about a hundred primary schools as an optional subject for pupils (see 4.4 for further details). The motivation for the introduction of these languages has been cultural rather than linguistic - the use of the vernaculars is encouraged as a gateway to Hindu and Muslim culture and religion, which many perceive to be on the decline under western influence. The effort is probably too late - it is not uncommon for an attempt to be made to bolster up and sustain an obsolescent language only after it has been eroded by another.

1.4 "Hindi"/Bhojpuri in South Africa: We now turn specifically to "Hindi" - the Indian language with the second largest number of speakers in South Africa. Like the other languages, it is markedly on the decline - a fact not adequately reflected by the following figures taken from various census reports for this century:

1936	1951	1960	1970
60 276	89 145	126 067	116 485

Although the instructions on the 1970 census form required the language most commonly used by each person at home to be recorded, the figures do not seem accurate. Taking the overall picture from the 1970 census, we have a population of 630 372 Indians, of whom 211 000 have English and/or Afrikaans as first language. The remaining 419 000 have some Indian vernacular as first language. Yet my contact with families suggests that even in 1970 the ratio of speakers with English as first language to the rest of the household is something in the region of 5 : 3. This hypothetical ratio is based on an average family's comprising a grandparent and two parents whose dominant language is an Indian one, and five children with varying degrees of proficiency in the vernacular, but for whom English (or Afrikaans in certain parts of the country), is clearly the first language. D. Bughwan's research (1970) on the use of English as mother tongue by Natal Indians supports this estimate.

An estimate of the number of speakers of an Indian language for 1970 would therefore be 3/8 x 630 000 ie. 236 250. The ratio of Hindi speakers to the total number of speakers of an Indian language, using the 1970 census data, is 116 : 416 . This ratio remains reliable, even though the individual figures are incorrect. (They all need to be proportionally downgraded.)

For 'Hindi' the number of L1 speakers can thus be calculated to be 116/416 x 236 250 ie. 65 877.

The figures for the early years (1936-1950) are, on the other hand, quite reliable since a vernacular was invariably the first language of the whole family. The figures for 1960 probably need to be down-graded by a few thousand.

The earliest "Hindi" - speaking immigrants came from various northern provinces of India, stated below in decreasing order of importance with respect to number of immigrants: Bihar, Uttar Pradesh (Agra and Oudh), Madhya Pradesh, Orissa, and West Bengal. The caste-affiliations of these people were various. Kuper (1960:272) cites the passenger list of the S.S. Congella, which arrived in Durban in 1889, as an example:

90	Camār	(low-caste leather workers)
39	Ahīr	(cow-herds)
30	Kahār	(fishermen, well-sinkers, domestics)
22	Kōlī	(cultivators and labourers)
15	Pāsī	(toddy-drawers)[6]
14	Dusadh	(corpse-bearers)
13	Bhur	(pig-rearers)
12	Kunbi	(cultivators)
11	Loniā	(salt-makers)
10	Chatrī	(warrior-rulers)
68	Other	(various, with fewer than members each - barbers, gardeners, shepherds, washermen, liquor sellers, etc.

Table 5 - Caste groupings of passengers on the SS. Congella (source: Kuper 1960).

The language spoken by these North Indians is termed "Hindi" by local historians and language teachers, but is not the term preferred in the writings of linguists. Gumperz and Naim, in a widely acclaimed article, discuss the rise of Hindi as follows:

> Modern Hindi-Urdu had its origin in the speech of the Gangetic Doab, just north and north-east of Delhi, in the districts of Meerut and Moradabad. Soon after the Muslim Conquest, a trade language, based on this speech, became the lingua franca of the courts, army camps and trading centres of the new ruling groups. The use of this language

> for trade purposes spread through much of North India, into many districts where the native dialects were quite different, as far as the coastal district of Bengal, Gujarat and Maharashtra. Some hundred years later we find records of several literary styles bearing various degrees of similarity to the spoken trade idiom. Among these are the Khari Boli of Delhi, the Dakini of the Muslim courts of the Bahami kingdoms in the Deccan ..... and several others ..... The development of the modern standard language did not really begin until the second quarter of the nineteenth century, after the consolidation of British rule, when literary Urdu replaced Persian as the official language for the purpose of local administration .....
>
> (Gumperz and Naim in Gumperz 1971: 52-53)

A new prose style in Urdu subsequently developed, gained prestige and came to be taught in schools. This variety was, however, not acceptable to all of North India, because of its association with Islamic culture and religion. A second prose style was developed by purging the language of its Persian element , as far as possible, and replacing it by borrowings from Sanskrit. This 'Hindi' (or 'Standard Hindi') gained currency among religious reform movements, schools and colleges. According to Gumperz and Naim, the situation by the end of the nineteenth century was that these two new standards, Hindi and Urdu, had almost replaced the sub-regional dialect literatures in the present Hindi area, but did not extend into places like Gujarat, Bengal and Bombay where other literary styles were already established, though spoken varieties of Hindi existed there as low-prestige trade idioms. After independence in 1947, Hindi replaced Urdu throughout the regional language area as official language of administration and education, though Urdu continues to flourish in literature.

Hindi, then, is a standard language whose native speakers represent an educated urban minority, though a growing one. Its presence is felt throughout India in newspapers, schools and universities, and over the radio, but its native speakers are spread widely over cities like Karachi, Bombay, Calcutta, Hyderabad, and the capital city, New Delhi. This relatively new standard differs greatly from the Indic languages and dialects spoken in rural North India which form a chain of mutually

intelligible varieties which vary geographically from village to village, and socially from caste group to caste group.

Superposed above this chain is, as Gumperz explains, a third variety, different from both colloquial language and Standard Hindi (but related to both), which he terms 'sub-regional dialects'. These are varieties which avoid the extreme localisms found in village speech and are consequently understood over much wider areas. They are the native languages of the castes which traditionally engage in trade in the small bazaar towns and larger city centres, and who cater for the needs of the rural population. These rural-dwellers themselves employ such sub-regional dialects as a second speech-style for trading purposes and for inter-group communication. The status accorded to these sub-regional and regional languages, by government and administration, is not consistent. Some are recorded as official languages of their state (for example, Panjabi in Panjab), while others are acknowledged as 'principal' languages, but are not officially sanctioned for governmental and administrative purposes (for example, Rajasthani).

We have distinguished, then, between 'Hindi' as a loose cover term (sometimes 'Hindustani')[7] for a whole range of varieties from Bengal to Panjab, and 'Hindi' (or 'Standard Hindi') as one particular variety - the literature-influenced speech of cities like Delhi, Bombay etc. Following Grierson's monumental Linguistic Survey of India, issued in eleven volumes between 1903 and 1928, most linguists today use the term 'Hindi' in the latter sense only, and prefer to think of the other varieties as separate languages with different names - Awadhi, Assamese, Rajasthani, Bihari etc.

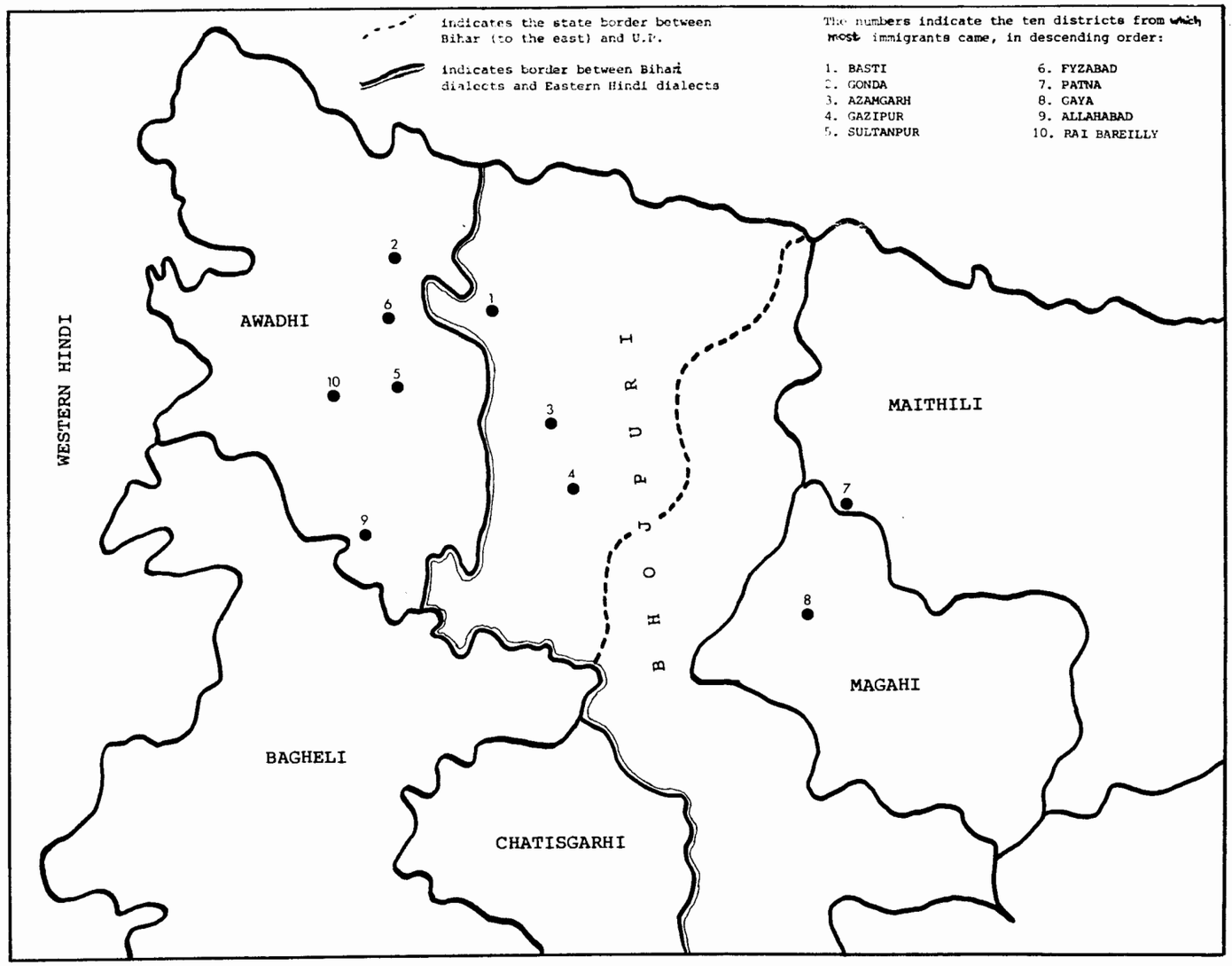
Grierson (1927: 158) draws a further distinction between 'Eastern' and 'Western' Hindi, treating them as two different languages. He uses the term 'Western Hindi' to denote the language whose dialects include Standard Hindi as well as the important Braj Bhaka[8], while 'Eastern 'Hindi' is used as a cover term for a group of dialects intermediate between 'Western Hindi' and languages further east like Bengali - of which Awadhi is the chief.

Our concern in this study is with the two main dialects brought over to South Africa - Bhojpuri and Awadhi - as well as languages represented by a small minority of speakers - Braj, Standard Hindi, Rajasthani etc. Bhojpuri is the principal dialect of Bihari, an Outer language in Grierson's terminology, while Awadhi is the chief dialect of the Intermediate language, Eastern Hindi.

We now attempt to determine the relative contributions of these various languages to the formation of South African "Hindi", using Maureen Tayal's record (1980:357) of the places of origin of a random sample of 1384 migrants who had embarked at Calcutta (adapted as table 8). (Since the total number of migrants during the period 1860 - 1911 from North India was approximately 6000, this sample seems quite reliable for our purposes.) While the ships' lists record details like place of origin, age, and identification marks of each migrant, no mention is made of home language. Assuming that the home language of an individual coincides with the predominant language of his home-village, and using the detailed linguistic maps given by Grierson in his Linguistic Survey of India (a work which coincides roughly with the period of indentureship under consideration), we arrive at the following picture (table 6):

Dialect	No of Speakers
Bhojpuri	500
Awadhi	447
Magahi	97
Kanauji	69
Bengali	47
Standard Hindi	43
Chatisgarhi	39
Maithili	36
Rajasthani	17
Braj	14
Bundheli	13
Nepali	7
Bagheli	6
Oriya	3
Panjabi	3
Other (indeterminate)	43
	1384

Table 6 - No of speakers of individual dialects among the original migrants (based on random sample).



Map 3 - Sketch Map of the languages of N.E. India
(based on Grierson 1903).

In terms of Grierson's classification four language groupings can be discerned here: Bihari, Eastern Hindi, Western Hindi, and 'other' languages. The table below classifies the dialects given in table 6 above in accordance with these language groups, and gives the respectives percentages of speakers of each language.[9]

Language	Breakdown by Dialect		
Bihari	Bhojpuri	500	
	Magahi	97	
	Maithili	36	
		633	(46%)
Eastern Hindi	Awadhi	447	
	Chatisgarhi	39	
	Bagheli	6	
		492	(36%)
Western Hindi	Kanauji	69	
	Std Hn	43	
	Braj	14	
	Bundheli	13	
		139	(10%)

Other Languages		
Bengali	47	(3%)
Rajasthani	17	(1%)
Nepali	7	(0,5%)
Oriya	3	(0,2%)
Panjabi	3	(0,2%)
	77	(5%)

Table 7 - No of speakers according to language groupings among the original migrants.

Basti	98	Hardoi	15	**Manipur**	3	Muzaffarnagar	1
Gonda	83	Purulia	13	Rewa	3	Ranchi	1
Azamgarh	69	Farrukhabad	11	Pilibhit	3	Sirauhi	1
Gazipur	66	Muzaffarpur	11	Ayodhya	2	Cuttack	1
Sultanpur	62	Alwar	10	Dattia	2	Jabbalpur	1
Fyzabad	55	Bareilly	9	Mymansing	2	Bardwan	1
Patna	52	Chapra	9	Ambala	2	Bundelkhand	1
Gaya	48	Hazaribagh	9	Badaun	2	Dulamow	1
Allahabad	48	Hamirpur	8	Bankura	2	Ganjam	1
Rai Bareilly	46	Dholepur State(?)	7	Bulandsahar	2	Indore	1
Lucknow	45	Nepal	7	Hoshiarpur	2	Lahore	1
Gorakpur	38	Tirhut	7	Kalpi	2	Moradabad	1
Jaunpur	37	Agra	6	Banda	2	Panipat	1
Partabgarh	37	Aligarh	6	Lodhiana	2	Rohtak	1
Raipur	37	Etawah	6	Jodhpur	2	Darbhanga	1
Banaras	28	Bharatpur	5	Orai	2	Midnapur	1
Arrah	25	Shahjehanpur	5	Surat	2	Bhopal	1
Barabanki	25	Gwalior	4	Tikamgarh	2	Batia	1
Monghyr	24	Sitapur	4	Chittagong	1	Dinapur	1
Bahraich	23	Etah(?)	4	Haripur	1	Hoogly	1
Shahabad	23	Bhagalpur	4	Karauli	1	Kishangarh	1
Jaipur	20	Chota Nagpur	4	Patiala	1	Maksudabad	1
Unao	19	Nawabgunj	3	Deoghar	1	Purnea	1
Cawnpur	18	Jhansi	3	Bhumihar State(?)	1	Rampur	1
Fatehpur	18	Mathura	3	Delhi	1	Sowuth	1
Mirzapur	16	Meerut	3	Girdaspur	1	Dumka	1
Saran	16	Ajmere	3	Kheri	1	Samtar	1
Ballia	15	Gurgaon	3	Kasbah	1	Songhat	1

Table 8 - Places of origin of a random sample of North Indian immigrants.

(source: Tayal 1980)

These figures should be considered as approximative rather than absolutely correct, for the following reasons:

a) They are based, as already mentioned, on a random survey of the ships' records, rather than the total number of North Indian immigrants. In addition Tayal's figures do not include the small number of migrants who came from Mauritius.

b) There is not always a simple correspondence between a 'linguistic area' (which is set up on the basis of the dominant language in it), and the language(s) spoken by any resident of that area. For example, whereas Fyzabad is a village located in a predominantly Awadhi - speaking area, in Eastern Fyzabad, Bhojpuri predominates (Grierson 1904:9). In such (rare) instances I have tried to calculate the numbers proportionally, without any illusions of being entirely accurate.

c) The existence of a small proportion of Muslims in the sample complicates the picture somewhat. Although the language of these migrants and their descendants is recorded as 'Urdu' I have not included that language in table 7 above, because it was impossible to determine from Swan's survey the number of 'Urdu' speakers. It is likely that most indentured Muslims spoke the language of the village they came from - mainly Awadhi and in some instances Bhojpuri. The "Urdu" of many Muslims in South Africa today, especially the colloquial variety of the less-educated, seems to me more akin to Awadhi than any other language/dialect. The figure for Awadhi in table IV may need to be downgraded somewhat to accommodate those who did speak Urdu.

Nevertheless it emerges quite clearly that the Bihari language is the one with the largest number of speakers, amongst the South Africa-bound North Indian languages, with the Bhojpuri dialect being the most common manifestation of that language, and that the Awadhi dialect of Eastern Hindi is next in importance in this study.

Of the languages given above, Maithili and Magahi are dialects of Bihari mutually intelligible with Bhojpuri, while Bagheli and Chatisgarhi are Eastern Hindi dialects not very different from Awadhi in their essential grammatical forms. Kanauji is a dialect of Western Hindi spoken in Western Uttar Pradesh and is closer to Braj Bhaka than to Standard Hindi. Rajasthani and Panjabi share more common features with Standard Hindi than with Bihari, while the opposite is true of Oriya and Bengali. Nepali, which Grierson classifies with Western Hindi, borders on all of Hindi, Awadhi and Bihari. Our figures show that Standard Hindi played little part as a vernacular for these migrants, although there must have been a small proportion of priests and teachers who were thoroughly competent in it. The vast majority of the migrants, however, were unschooled and unacquainted with educated varieties of Hindi. Gumperz's remarks on the language problems in rural North India (1971:21) are illuminating:

> Lectures on technical subjects ... (in Standard Hindi)... by outside officials are also likely to present difficulty. The village-level worker and the officials from the local development centre who have been in the area for a while have no trouble in making themselves understood. There have been some lectures, however, by outside technical experts, notably one on artificial insemination, which were understood by only a few of the literate villagers. There are five or six radios in the village, but people do not ordinarily listen to the All India news broadcasts because they say the language is too difficult for them.

The relationship between Standard Hindi and the two principal Northern dialects brought to South Africa, Bhojpuri and Awadhi, is not at all understood by speakers of these dialects, nor by most teachers of Hindi in South Africa. Most people, learned and lay alike, characterise "South African Hindi" as a debased, "broken" language fit only for the "kitchen" - a distorted notion stemming from invalid comparisons with the standardised Hindi of Indian film, radio, and formal education.

The Bihari language (really a linguist's construct, since people claim one of its three dialects - Bhojpuri, Maithili, and Magahi to be their native language and do not use the term Bihari) had according to the

1970 census of India 16 806 691 native speakers in that year, making it the tenth most widely spoken language of the country. This figure needs to be increased by several millions to gain a truer picture, since many who receive their education in Standard Hindi claim that language as mother tongue instead of Bihari. Writing in 1903, Grierson indicates the number of Bhojpuri speakers alone to be twenty million, while Shukla (1968:ix) states that Bhojpuri had thirty million speakers in 1968. One scholar characterises the function of Bhojpuri as follows:

> Bhojpuri is a language which is very much alive. Although the primary and secondary education in the Bhojpuri area is imparted through the medium of the Standard Hindi and Urdu and the literary language is Hindi and Urdu, yet Bhojpuri occupies a place of honour and prestige in the hearts of its speakers. The oral explanation of difficult portions in Hindi and Urdu is frequently made in class in Bhojpuri when teachers and students both are Bhojpuri speakers. The students, both in their classrooms and outside, talk to each other in Bhojpuri and they would even address the teacher in the mother tongue in the lower classes. (Tiwari 1960:xxx)

The name Bhojpuri is derived from Bhojpūr, once an important old town in western Bihar. Bhoj denotes the warrior clan of rulers of the city in earlier times - the Ujjaini Bhojas, while pūr is the word for 'town' common throughout the country. In time, when the fame of this clan spread, the entire area to the south took its name from the town, while the term Bhojpuri gradually came to refer to the people as well as the speech of the surrounding areas, on account of the fame of the Bhojpur Rajputs (or warrior-rulers) in Moghul times. (Tiwari 1960:xxiii). The Bhojpuri area today covers most of North Bihar as well as Eastern Uttar Pradesh.

Bhojpuri does not have a great literary tradition, unlike its sister dialects, Maithili, which has an established literature, and Magahi, once prestigious as the language variety favoured by the Buddha and his followers. However, many folk-songs, poems, short stories, and folk-dramas are often written down and are becoming increasingly popular in print. The script employed in recording Bhojpuri is the Kaithi, an

adaptation of the Devanāgarī, also used extensively in the state of Gujarat, though more recently writers and printers are making increased use of the Devanāgarī script, to bring their literature in line with Hindi.

The Bihari dialects are, as already stated, more closely related to Bengali, Assumese, and Oriya than to Western Hindi, a linguistic fact which does not accord with socio-geographical reality, since Bihar has historical ties with the northern Indian province of Uttar Pradesh rather than with Bengal. All family ties and traditions point to the West, not the East where Bengal lies. Accordingly in the folk-linguistics of the people, Bihari is misconceived as being a substandard variety of (Western) Hindi. Some of the features common to the grouping Bihari-Bengali-Oriya-Assamese which are mistakenly cited as "substandard" Hindi are discussed below:

a) the past tense in -l, versus -ā or -ī in Hindi. eg. Bhoj: ham gai-lī Bengali: āmi ge-lām, Oriya: mũ ga-li, but Hindi: mẽ ga-yā (masc) 'I went'.

b) the future tense in -b, against -gā in Hindi. eg. Bhoj: ham dēkh-ab, Bengali: ami dekh-iba, Oriya: mũ dekh-ibi, but Hindi: mẽ dekh-ũgā (masc) 'I will see'

c) Adjectives are not declined for gender and number, whereas a large class of Hindi adjectives do take these endings: eg. Hindi: mīṭhā santrā 'sweet orange' (masc), mīṭhī cay 'sweet tea' (fem) mīṭhe santre 'sweet oranges' Bhoj uses the invariant adj. mīṭhā in these instances, except for some Hindi-influenced Western sub-dialects.

d) Unlike Hindi, the eastern group employs highly irregular oblique forms in the declension of nouns and pronouns: eg. Bhoj: hamani-ke Maithili: ham-ār Bengali: āmār, but Hindi: ham 'us' (oblique form) (For other examples see Grierson 1903: 2-3).

The other dialect featuring prominently in the language history of Natal, Awadhi, is grouped by Grierson together with Bagheli and Chhatisgarhi under the cover term 'Eastern Hindi', an intermediate variety between Western Hindi and the Eastern group of Bihari, Bengali, Oriya etc. The term 'Awadhi' denotes the language of Awadh (or Oudh), one of the old provinces of the north, which together with the former province of Agra is today known as Uttar Pradesh ('North Province'). Awadhi is spoken throughout central Uttar Pradesh, with the exception of a few districts in which Bhojpuri prevails. In addition Awadhi is the vernacular language of many Muslims living in areas where a Bihari dialect predominates. Oudh is famous for its ancient capital Ayodhya - birth-place of the hero Rama, whose exploits are celebrated in epic literature and folk-songs, much of which was composed in the language of the area. Tulasi Das wrote his version of the Rāmāyaṇa in Awadhi in the late sixteenth century. Although Awadhi is not an official state language today, it is still looked upon as a natural vehicle of epic verse, including - among others - the standard vernacular translation of the Sanskrit Mahābhārata.

We conclude this section with a brief discussion of the sub-dialects of Bhojpuri. Bhojpuri, covering some fifty thousand square miles, has four main sub-dialects - the Standard, Western Bhojpuri, Nagpuria, and Tharu (Grierson 1903:42). Using Tayal's table of the places of origin of North Indians, we arrive at the following figures for the number of speakers of individual sub-dialects of Bhojpuri arriving in South Africa:

North Standard	39%	56%
South Standard	17%	
Western	44%	
Other	0,4%	

I shall use the term Eastern Bhojpuri instead of Standard since the latter term has never been well motivated by any writer. There is till

today no single accepted standard Bhojpuri.

The main differences between the Eastern and Western dialects are the following (after Grierson 1903):

1) The postposition 'for, of' is in Western Bhojpuri ka or kai with an oblique alternative in ke. In Eastern Bhojpuri the base form is ke, the oblique ka.

2) Western Bhojpuri has an instrumental case in -an, lacking in the Eastern variety.

3) Whereas the Western dialect sometimes inflects adjectives for gender and number, the Eastern does not.

4) The honorific second person plural pronoun is tũ in the Western dialect, raure in Eastern Bhojpuri.

5) The oblique marker for nouns and prounouns is -e in the Western dialect, -a in the Eastern.

6) The first person present tense form of the verb 'to be' is hauwi in Western Bhojpuri, howi in the Eastern.

7) Whereas the first person singular and plural verb forms end in ĩ in the Standard dialect, nasality is lacking in the plural form in the Western dialect, while occasionally remaining in the singular.

8) The third person plural, present tense of verbs ends in ai in Western Bhojpuri, the Standard ending is -an or -ani.

The differences between the North and South Eastern varieties are minimal, the two chief ones listed by Grierson are that Northern ṭ becomes ṛ in verb paradigms in the South and that Southern l̤ becomes

u in the past tense of verbs in the North.

Awadhi, on the other hand, is spoken more or less uniformly over a large area, with only minor variations from district to district.

NOTES

1. Kuper (1960:7) gives the following figures: "Among 3 200 indentured coming on 8 boatloads selected at random, approximately 2% were Brahmin, 9% Kshatriya, 21% Vaishya, 31% Sudra, 27% scheduled castes. Of the remaining groups 3% were Christian, 4% Muslim and 3% unclassifiable."

2. These totals are based on figures in India - A Reference Annual 1973, and reflect the language groupings of the 1961 Census.

3. Note that Emeneau (1956) specifically excludes the Tibeto-Burmese family from this characterisation of the Indian linguistic area, since it does not share a large core of common features with the other three groups.

4. The exposition is based primarily on Grierson (1927), though additional discussions by Chatterji (1960), Burrow (1955), Zograph (1982) and Shapiro and Schiffman (1981) have been useful.

5. The 1911 Census figures are given as follows:

Total Indian Population: 149 791 South African-born: 63 776

Remainder (Place of Birth):

Assam	31	East Bengal	3
Bengal	16 165	Madras	41 314
Bombay	10 883	Punjab	342
Burma	33	U.P. (Agra and Oudh)	265
Central Provinces and Bihar	49	Other & Unspecified	15 921
		Other Countries	1 009

These figures are incorrect. The number given for immigrants from Bihar is much too low, while the figures for Bengal are too high. It seems that respondents reported port of embarkation (Calcutta, which is in Bengal), rather than birthplace.

6. A toddy-drawer is a worker who taps palm trees for the sap (toddy) used in making liquor. The term pāsī also refers to a fowler.

7. The term 'Hindustani', avoided in this study, is used in at least three different senses by writers:

   a) In British India the term was interchangeable with 'Hindi' - denoting the regional language of parts of North India (Uttar Pradesh, Madhya Pradesh), as well as the language used in larger cities like Bombay.

   b) 'Hindustani' is used by others to denote the colloquial variety (devoid of learned Sanskritic borrowings) associated with the less-educated. The term is used in opposition to formal Hindi.

   c) It is sometimes used as a cover term for both Hindi and Urdu which are seen as two variations of a common-core language, the latter having a heavily Persianised lexicon, while the former leans more towards Sanskrit in its lexicon.

8. Braj (or Braj Bhākhā), spoken in Western Uttar Pradesh, has a long tradition of music and poetry behind it. It is especially linked with poetry written in praise of the God Krishna. Grierson (1916:72) claims this dialect to be more typical of Western Hindi and more archaic than literary Hindi.

9. The percentages are calculated proportional to the total number whose language could be ascertained using Tayal's random survey. The total is 1349, which excludes 43 whose home language could not be ascertained in this way.

CHAPTER 2

A SKELETON GRAMMAR OF INDIAN BHOJPURI

In this chapter the main characteristics of contemporary Indian Bhoj day are presented in outline. The discussion is, of necessity, brief, and therefore incomplete in many respects. It is not the intention to treat Bhoj grammar exhaustively for its own sake, but to provide an overview from which the internal history of Bhojpuri in South Africa can best be gauged in subsequent chapters. The reader seeking further details should consult Shukla (1968, 1983), Tiwari (1960), Misra (1980), and Grierson (1903, 1883-1887), to all of whom I am indebted in this chapter, and whose judgements I have attempted to corroborate by field-work in western Bihar.

Some of the difficulties facing a grammarian of Indian Bhoj are the immense regional variation, the absence of a standard form of the language, and the constant influence from closely related languages/dialects like Awadhi in the west, Magahi and Maithili in the east, and more importantly, Standard and regional Hindi. Although previous writers have used such terms as 'Standard North', 'South Standard', and even 'Standard Bhojpuri' (based rather uncritically on Grierson's classification), no such single or multiple standard of speech or writing has yet evolved. In this chapter I attempt to present an overall picture of the features which are common to most sub-varieties, rather than focussing on one dialect alone. There is accordingly some variation, mainly morphological, in the data : for example, I sometimes use the copula form characteristic of Eastern Bhoj bā (and its variants), while at other times preference is given to the form which is an alternative in Western Bhoj - hai.

2.1 Phonetics and Phonology:

2.1.1 A typological overview: From the vantage-point of the phonological

typology of Indic languages, Bhojpuri is unremarkable - having virtually no features which are not found in almost all the others. Ramanujan and Masica (1969), on the basis of certain phonological criteria, drew typological maps of the languages of all of India, which are useful for our purposes of placing Bhoj within the Indic family. These criteria are:

a) Retroflex consonants: Retroflex phonemes occur in all languages of the Indian sub-continent, except in the North-East languages (eg. Assamese) and Baluchi in the north-west. Bhoj has retroflex phonemes among stops only, as do Hn, Bengali, Nepali, and Kashmiri, while Gujarati, Marathi, Konkani, and Oriya have in addition phonemic /ɭ/ and /ɳ/.[1] The North-West group of Lahnda, Marwari, and Panjabi have retroflex stops as well as phonemic /ɭ/, /ɳ/, and /ɽ/. (/ɽ/ exists in most Indic languages, but only as an allophone of /ɖ/.) The Dardic languages of the north-west (which fall midway between the Indic and Iranian languages) like Kashmiri and Burushaski have retroflex stops and fricatives.

b) Aspiration and Murmur: All the Indic languages, except Sinhalese, have a series of phonemic voiceless aspirates. In addition all, except Panjabi, Sinhalese, and the Dardic languages, have a parallel series of murmured consonants (or breathy voiced consonants). In addition some languages have murmured nasals (eg. Rajasthani, Marathi, and Bihari), murmured liquids (Marathi, Oriya, Bihari), while only Marathi has murmured semivowels.

c) Nasal Vowels: All the Indic languages, except Marathi, have a phonemic contrast between nasal and oral vowels, though not necessarily for every vowel. Bhoj has a full set of long and short nasal vowels, corresponding to the oral set.

d) The /ts/ - /tʃ/ contrast: The North-Western group of Dardic languages (eg. Kashmiri and Khowar), Iranian languages (eg. Pashto), and Pahari have a dental (or prepalatal) affricate /ts/ as opposed to the palatal affricate /tʃ/, a contrast which is not found in Bhoj or most other Indic languages. In the South-Western languages (Marathi and Konkani) the relationship between the two sounds is an allophonic one.

e) /n/ versus /ɲ/: A contrast between the dental or alveolar /n/ and a palatal /ɲ/ occurs in some of the Indic languages of the north-east, including Nepali, and Maithili, though the majority of the Indic languages (including Bengali, Panjabi, Marathi, Hindi, and Gujarati) have [ɲ] as an allophone of /n/ in (alveo)palatal environments. There is some doubt concerning the status of /ɲ/ in Bhoj. Ramanujan and Masica, following Tiwari (1960), include Bhoj in the group of languages having phonemic /ɲ/, though Shukla (1983) and Misra (1980) consider it to be purely allophonic. I shall argue shortly that the latter position is the correct one.

f) Vowel Systems: Together with Hindi, Sindhi, Panjabi, Lahnda and other northern Indic languages, Bhoj has a five-vowel system, with length of vowels being phonemic. Like these languages Bhoj has a high-mid-low contrast, and for the non-low vowels a front-back contrast. In Marathi and Gujarati distinctions of length are lacking for some vowels, while further east in Bengali and Assamese only mid-vowels make a phonemic contrast of length.

2.1.2 The Consonants of Bhoj: Table 9 lists the consonantal phonemes of Indian Bhoj, while some of their main allophones are enclosed in brackets.[1] The retroflex flaps [ɽ] and [ɽʰ] are allophones of [ɖ] and [ɖʰ] respectively, occurring intervocalically and word finally after vowels. In Eastern Bhoj [ɽ] and [ɽʰ] are replaced in some words by the weaker alveolar flaps [ɾ] and [ɾʰ] respectively eg. W Bhoj [gu:ɽ] 'molasses', [gʰo:ɽɑ:] 'horse' and [ko:ɽʰi] 'leper' have the E Bhoj equivalents [gu:ɾ], [gʰo:ɽɑ:] or [gʰo:ɾɑ:] and [ko:ɾʰi] respectively.

The status of [ɲ] is, as already mentioned, in dispute. Tiwari cites the forms [niɲɲɑ:] 'sleep', and [bʰuiɲɑ:] 'earth' in support of his claim that it is a phoneme. He also emphasises that this [ɲ], resembling nasalised [j], differs from the nasal allophone occurring in the environment of (alveo)palatal affricates. Other writers like Shukla and Misra regard [ɲ] to be an allophone of /n/, occurring with /tʃ/ and /dʒ/, but remain silent on the words cited by Tiwari. It seems to me that the [ɲ] that Tiwari posits for 'sleep' and 'earth' is really [j̃], a nasalised allophone of [j], occurring between a nasalised vowel and a low back vowel.

		Bilabial	Dental	Alveolar	Retroflex	Palato-Alveolar	Palatal	Velar	Glottal
STOPS (Voiceless)	Non-asp.	p	t̪		ʈ			k	
	Asp.	pʰ	t̪ʰ		ʈʰ			kʰ	
STOPS	Voiced	b	d̪		ɖ			g	
	Murmured	bʱ	d̪ʱ		ɖʱ			gʰ	
NASALS	Voiced	m		n	(ɳ)		(ɲ)	ŋ	
	Murmured	mʱ		nʱ				ŋʱ	
AFFRIC. (Voiceless)	Non-asp.					tʃ			
	Asp.					tʃʰ			
AFFRIC.	Voiced					dʒ			
	Murmured					dʒʱ			
FRICAT. (Voiceless)	Non-Asp.			s					
FLAPS	Voiced			ɾ	(ɽ)				
	Murmured			ɾʱ	(ɽʱ)				
APPROX.	Voiced	w		l			j		
	Murmured			lʱ					ɦ

Table 9 - The Consonants of Bhojpuri

That this is not an ad hoc procedure can be seen from Tiwari's own characterisation of it as resembling nasalised [j] (1960:12), and from the fact that [w] shows similar nasalisation between a nasal vowel and -a, as in [$d^{h}\tilde{u}:\tilde{w}ə:$] 'smoke', and [$dʒ\tilde{a}:\tilde{w}ə:$] 'burnt brick'. [ɲ] in Bhoj can therefore considered an assimilatory allophone of /n/ adjacent to the palato-alveolar [tʃ] and [dʒ] as in [paɲtʃ] 'public', [gaɲdʒ] 'a bald person'.

/n/ has a further allophone, [ɳ] before retroflex consonants, and following a retroflex consonant plus vowel eg [t^{h}aɳɖa:] 'cold' (adj), [g^{h}o:ɽaɳ] 'horses'.

Although most traditional grammars of Bhoj and other Indic languages classify the voiced series (/b/, /d/, /ɖ/, /dʒ/, and /g/) as unaspirated, in contrast to a parallel group of "voiced aspirates", I have avoided this terminology since the series (/$b^{ɦ}$/, /$d^{ɦ}$/, /$ɖ^{ɦ}$/, /$dʒ^{ɦ}$/, and /$g^{ɦ}$/) are neither truly voiced (in the sense of being produced by having the vocal cords close together and vibrating), nor aspirated (which would imply a period of voicelessness after the articulation of the stop, and before the onset of a following vowel caused by the (aperiodic) expulsion of air). Instead they are best characterised as 'murmured' (or 'breathy-voiced'), ie. produced by vibrations while the vocal cords are apart and by a high rate of (periodic) airflow through the glottis, with no ensuing period of voicelessness (see Ladefoged 1982: 47-48 and 128-129).

Likewise Bhoj has phonemic murmured nasals [$m^{ɦ}$] , [$n^{ɦ}$], and [$ɲ^{ɦ}$] (with contrasts like ban 'arrow', versus banh 'embankment'), though neither the [ɲ] nor the [ɳ] allophone has a murmured counterpart. [$ɾ^{ɦ}$] and [$l^{ɦ}$] too are murmured rather than aspirated consonants, contrasing with [ɾ] and [l] in pairs like kōlā 'a small field' and kōlhā 'string of a spinning wheel', and mar 'to beat' and marh (a type of grain). None of these five murmured sonorants occur in initial position, nor is their total number very large.

Other minor consonantal allophones are fronted versions of the velar stops before front vowels, and the weakly trilled [r] in final position in place of the more usual weak alveolar flap [ɾ] in other positions.

The nasalised allophones of /j/ and /w/ have already been described. /ɦ/ tends to become voiceless in final position.

2.1.3 The Vowel System: There is much uncertainty in the literature concerning the vowels of Bhoj. Tiwari's characterisation (1960) is lacking in rigour, and sometimes contradictory; Misra (1980) lists five short oral vowels, and three long oral vowels, in addition to the nasal vowels; while Shukla (1983) posits six short oral vowels, five long oral vowels, and ten nasal vowels, equally divided into long and short sub-classes. Finally Trammel, describing - like Shukla - the 'North Standard Bhojpuri' lists five short oral vowels, six long oral vowels and an equivalent set of nasalised vowels.

I present what seems to be a 'common core' phonemic inventory of the vowels of Bhoj based on fieldwork in the Western Bhoj area, as well as on my interpretation of the above sources, followed by an outline of the main allophones and some dialectal differences.

ORAL VOWELS

	Front		Central		Back	
High	i	i:			u	u:
Mid	e	e:	(ə)	(ɐ:)	o	o:
Low					ɑ	ɑ:

NASAL VOWELS

	Front		Central	Back	
High	ĩ	ĩ:		ũ	ũ:
Mid	ẽ	ẽ:		õ	õ:
Low				ɑ̃	ɑ̃:

Table 10 - The Vowels of Bhojpuri

A few examples which establish the nasal vowels as phonemes are: j̃āt 'to press', versus jāt 'caste'; g̃ōṛ 'foot',versus gōṛ (a caste name); s̃ūngh 'to smell', versus sūn 'quiet', and so on.

The long vowels (oral and nasal) all have half-lengthened allophones at the end of words. /i/ has a lowered allophone [ɪ] usually in checked syllables eg [pɪllɑ̈] 'puppy', in at least some dialects (Trammel 1971). Although /e:/ is phonetically always [e:], /e/ is more often realised as /ɛ/, especially in checked syllables, though there seems to be free variation between [e] and [ɛ], eg. [belə wɑ̈:] or [bɛlə wɑ̈:] 'woodapple' (Trammel 1971).

Bhoj [o:] is considerably rounded, much more so than its short equivalent [o]. Although all the writers mentioned above consider Bhoj [ɑ:] and [ɑ] to be low central vowels, it seems to me [ɑ:] is most certainly a back vowel, while [ɑ] has a tendency to become centralised. In educated, Hindi-influenced Bhoj, /ɑ/ tends to be pronounced as [ə], even in monosyllabic words - eg. [mər] 'to die', [bəs] 'enough', and so on.

[ə] is a frequently occurring allophone of [ɑ], usually in the least stressed syllable of tri-syllablic words, usually formed by affixation: eg [gʰo:ṛa:] 'horse', [gʰoṛəwa:] 'the horse'; [rəɦɔl] 'she was', [rəɦəli:] 'I was'. In these two instances [ə] is voiced, though it is more frequently devoiced (in the more usual environment of a preceeding voiceless consonant - eg [dɛkʰə̥lɑs] 'he saw').

Eastern Bhoj has, in addition, two extremely shortened, voiceless vowels [i̥] and [u̥], in word final position - reflexes of final OIA i, ī, u or ū - which are lost in most Indic dialects. Some examples are: E. Bhoj [sɑ:su̥] 'mother-in-law', W. Bhoj [sɑ:s]; E. Bhoj [ri:ti̥] 'custom', W. Bhoj [ri:t] etc.

Another unusual feature (for Indic) of Bhoj phonetics is the mid, central, rounded vowel [ɵ], which occurs solely in verb paradigms. It is usually a marker of the second person neutral form, neither an honorific nor a (-R) form, which have an ī and -ē ending respectively. It does not seem

to occur in all dialects of Bhoj (and is not mentioned in Shukla's grammar of North-Eastern Bhoj) but is a striking feature of those varieties that do use it.

The presence of [æ] (or [æ:], depending on which grammarian's authority we rely) in Bhoj is rare, and, I think, characteristic of educated speakers who show the influence of Std Hn.

A significant phonological rule of Bhoj is that of the 'short antepenultimate' (Grierson 1903:24), which determines that a long root vowel is shortened if it becomes, by affixation, part of the antepenultimate syllable (or - rarely - further removed from the last syllable). Alternations of the following type are very common in Bhoj (and many other Indic languages): [si:kʰɑb] 'I will learn', as against [sikʰəbe] 'You will learn'; [ɑ:pɑn] 'oneself', as against [ɑpəne] 'by oneself', and [pɑ:ni:] 'water', as against [pɑniyɑ] 'the water'.

2.1.4 Syllables and stress: The canonical pattern for Bhoj syllables is (C) V (C) (C) - that is, a syllable may consist of a single vowel (eg. ū 'that'), a consonant plus vowel (eg bā 'he is'), a vowel plus one or two consonants (eg. ēk 'one', ānṭ 'end'), or a consonant plus vowel plus one or two consonants (eg bandh 'closed'). It is rare, however to have a syllable consisting of VCC alone. When two consonants occur medially, they are usually separated by a syllable boundary (for further discussion see Shukla 1983:31-45). Bhoj does not permit initial clusters, with words having a cluster in other dialects showing epenthesis eg Hn prān 'spirit', briyānī (a type of food), Bhoj parān and biryāni. Consonant clusters, where they do occur, may consist of geminates (eg ann 'grain'), unaspirated consonants followed by their aspirated counterparts (eg. sukkhal 'dry') (though these could be considered to be geminated aspirates with the aspiration fully realised at the end of the cluster), or nasal plus homorganic consonant (eg. gandh 'smell').

Stress is non-phonemic in Bhoj. The following set of rules account for primary-stress patterns in the language at a "surface" phonetic level (for further details, including the rules for secondary stress the reader is once again referred to Shukla 1983: 44-46):

a) The first long vowel in a word is stressed fully. By the rule of the 'short antepenult', outlined above, long vowels are found almost exclusively in final or penultimate position.

b) If a word has no long vowels, then either i) the first heavy syllable (containing at least a long vowel or diphthong, or CVC(C)) is stressed or ii) if there are no heavy syllables, the penult is stressed.

In the following illustrative examples, a syllable boundary is denoted by '-': phál 'fruit' (monosyllabic content words are all stressed), pā́-gal 'mad', dhaḍ-dhā́ 'langur', pī́-yā 'husband, lover', bhin-sa-ha-rā́ 'dawn' (all conforming to rule (a)); cál-al 'he went', ja-nám 'birth', lók-an-i 'people', la-ja-hár 'a shy person' (all by rule (bi)), and mú-ni 'a sage', sa-mú-ji 'understanding', and ba-hí-ni 'sister' (all by rule (bii)).

Words like dhaḍ-dhā́ and bhin-sa-ha-rā́ show that syllables with long vowels have priority over other heavy syllables in the stress rule.

2.2 The Bhojpuri Lexicon: The traditions of Indian grammarians of the past still provide the most useful terminology and framework for the descriptions of the lexicon of any contemporary Indic language.

2.2.1 Tatsamas, Tadbhavas, and Deśya words: The term 'tadbhava' refers to words belonging to NIA languages which are traceable to Sanskrit and the Prakrits, via the Apabhraṁśas, and which show the phonetic changes characteristic of the MIA stage (notably the simplification of consonant clusters, the use of geminates, the frequent change of aspirated and murmured consonants in medial position to an h, the nasalisat-

ion of vowels, etc). Tadbhava means 'derived from that' (ie. from Sanskrit), although we now acknowledge that Sanskrit was not necessarily the direct source for these words. The words of this group are fundamental ones in all the Indic languages of today, and the following tadbhavas of Bhoj are found in virtually the same form in the other Indic languages: hai- 'to be', kar- 'to do', ā- 'to come', jā- 'to go', khā- 'to eat', pī- 'to drink', mar- 'to die', mār- 'to beat', sūn- 'to listen', dēkh- 'to see', tū 'you', hāt 'hand', nā̃k 'nose', ā̃kh 'eye', dā̃t 'tooth', suruj 'sun', tāra 'star', gāi 'cow', and many more. Many of the tadbhavas are, of course, of Indo-European origin, and typically show considerable changes from their Indo-European and OIA prototypes. For example Bhoj sā̃p 'snake', bhītar 'inside', bhīj 'to drench', and ī̃ṭ 'brick' are reflexes of very different OIA forms, sarpa, abhyantara, abhyañj, and iṣṭa respectively (derivations from Tiwari 1960).

Tatsamas are borrowings from Sanskrit by Indic languages in modern times, which retain some Sanskritic phonetic combinations which had undergone change in MIA. Tatsama means 'same as that' (ie. as Sanskrit), though modern pronunciations of tatsamas do differ from the Sanskrit originals. The word rājā 'king' in Bhoj and many Indic languages is a tatsama, existing side by side with the tadbhava form rao, and the Bhoj honorific pronoun rauwa. Tatsamas are usually treated as borrowings in Indic languages, by not taking on inflections - eg. rājā in Hindi does not have the usual oblique ending in -e.

Whereas literary languages like Hindi and Bengali were at one stage characterised by a high degree of Sanskritisation, tatsamas in non-literary languages like Bhoj are few, and have entered the language via the influence of Hindi, rather than directly from Sanskrit. Some tatsamas used in Bhoj by the more educated are ānand 'joy' (which might in less-educated speech occur as ānan), rājanīti 'politics', buddhi 'intellect', vidyārthī 'student', and prabhaw 'influence'. Sanskritised names for deities like Krishna (or Kṛṣṇa), Shiva, and Vishnu have their counterparts in the tadbhavas Kisūn or Kānhā, Siū and Bisūn.

The term 'semi-tatsama' was introduced by Western grammarians for those old literary borrowings from Sanskrit in MIA, which then show subsequent changes undergone by other words in a particular vernacular. An example given for Hindi by Grierson (1927:128), but which applies equally to Bhoj, is the development of Classical Sanskrit vaṁśa which could mean either 'family' or 'bamboo' : the semi-tatsama form bans means 'family' while the tadbhava form bā̃s, showing nasalisation of vowels upon loss of a pre-consonantal nasal, has the meaning 'bamboo'.

Deśya words are those of non-Indic origin which are to be found in MIA, and the vernaculars. These include borrowings from Dravidian and Munda sources (deśya = 'of the country, local'). Bhoj, like most of the modern Indic languages, has many onomatopoeic words which are probably of Dravidian origin (eg. phū̃k 'to blow', suruk 'to inhale', and kanmanā 'to murmur'), and other words like pagaṛī 'turban', khaṭṭa 'sour', gāṛ 'to bury', and ṭikkā 'a dot on the forehead'.

Of course Sanskrit had already absorbed some Dravidian and Munda loans prior to the MIA period; but these were not recognised as deśya words by early grammarians, who also erred in sometimes categorising words of Indic origin that survived in Prakrit, but not Sanskrit, as belonging to the deśya category.

2.2.2 More recent borrowings: a) Perso-Arabicisms: The conquest of India by several Muslim powers - Turks, Persians, Mughals - spanned many many centuries, culminating in the Mughal Empire in the sixteenth century; this brought several languages to India - among them Persian, Arabic, and Turkic languages. Of particular importance in the history of NIA is the influence of literary Persian (itself owing much to Arabic), cultivated by the Moghul administration, which resulted in the absorption of a high percentage of loans into Urdu/Hindi, and subsequently into other neighbouring languages. Some of the Bhoj words traceable to Persian (and eventually to Arabic) in this way are: mālik 'Lord, master', namāj 'Muslim prayers', kāgaj 'paper', gōs 'meat', tasbīr 'picture', nagīc 'near', gujar- 'to pass away', badal- 'to change (one's

clothes)', kamti 'a little', khus 'happy', rumāl 'handkerchief', jaldi 'quickly' and others.

b) Europeanisms: Direct borrowings in the NIA period go back to the fifteenth and sixteenth century to Portuguese, Dutch, and French contacts with India, though not many Bhoj words have a continental European origin - mistrī 'artisan', brandā 'verandah', and kamrā 'room', for example, are all traceable to Portuguese.

The British rule in India led to an exchange of loanwords between English and many languages on the sub-continent. Although Bhoj shows little of the predilection for Anglicisms that Std Hn has, a number of English words are now in constant use, often entering the language via Bengali or Hindi. To this group belong ṭikaṭh 'ticket', ḍiptī 'deputy', heḍ-māhṭar 'head-master', aksiḍanṭ 'accident', aleksan 'election', mōṭar 'car', aparēsan 'operation', laībareri 'library', ṭaim 'time', bank 'bank (financial)', āphis 'office', dāgḍar 'doctor', ispīṭ 'speed', iskūl 'school', hoṭal 'hotel', bel-bāṭam 'bell-bottomed trousers' etc.

c) Borrowings from other Indian languages: In addition to many borrowings from Hindi, the official language of both states in which Bhoj is widely spoken, which are themselves often traceable to various foreign sources, Bhoj shows the influence of other NIA languages, especially Bengali. From Bengali come such words as rasgulla (a round sweetmeat), murhī 'fried rice', bāsā 'house', phālī 'piece', and a few more domestic terms.

2.2.3 Characteristic Lexical Patterns: To conclude this brief characterisation of the Bhoj lexicon, we examine patterns of compounding and reduplication which are quite extensive in the language.

a) Compounds: In describing compounding, the terminology of Sanskrit grammarians, once again, proves useful. The data for Bhoj is taken from Tiwari (1960).

i) Dvandva Compounds: These involve combinations of co-ordinate nouns, the term dvandva meaning 'two by two'. In the following examples the literal meaning of each compound is given, followed by a more idiomatic gloss within brackets: māī-bāp 'mother-father' (ie. 'mother and father' or 'parents'), hāth-gōṛ 'hand-foot' ('limbs'), rāt-din 'day-night' ('night and day'), dāl-bhāt 'dal (or split lentil)-rice' ('a meal of dāl and rice').

In some instances more than two nouns may be conjoined: hāth-gōṛ-nāk-kān 'hands-feet-nose-ears', ('bodily appendages'), dāl-bhāt-sāg 'dal-rice-vegetables' ('a meal of dal, rice, and vegetables') etc.

A related type of compound uses rhyming or alliterative synonyms: kām-kāj 'work-work', māth-mūr 'head-head', ghar-bārī 'house-house'. These may function as emphatic equivalents of single nouns ('work', 'head' etc) or have the collective sense of 'work and such-like', 'head and other parts of the body' etc.

ii) Determinative Compounds: To this category belong compounds having one member (usually the first), qualifying the other. Further subdivisions of this catgory are: dwigu (or numeral determinatives) eg cau-mukh 'four-faces' ('facing four sides'), nawa-ratan 'nine-jewels' ('having nine jewels') etc; tatpuruśa (subordinate determinative compounds)[2] the most important examples in Bhoj are those having a noun whose (deep) case is understood to be an oblique one, even though the postposition expressive of such cases is absent eg bijulī-māral 'lightning-struck' (ie 'struck by lightning' with the instrumental postposition se understood), Gangā-jal 'Ganges-water' (ie. 'water of the Ganges', with the genitive postposition ke understood), jal-khaī 'breakfast-eating' (ie 'the act of eating breakfast', with the first noun understood to be accusative); while a third subdivision consists of other appositional determinatives eg mahā-rānī 'great-queen', kāc-kēlā 'green-plantain', etc.

iii) Bahuvrīhī Compounds: These are compounds which involve predications of a third party, eg. lāl-pagarī 'policeman' (literally 'red-turban' - so the compound refers to one who wears a red turban, not to a turban which

is red), rukh-caṛhawā 'monkey' (lit. 'tree-climber'), and ghaṭ-phoṛawā 'brahmin' (lit. 'earthern pot breaker').

b) The Echo-Word Construction: This is a special type of compound occurring in many languages of India (see 1.2), in which a word is duplicated, with the first syllable being systematically changed. This type of compound, which conveys a sense of 'collective-ness', differs from others in that the second element (the 'echo') does not have independent status in the language. Some Bhoj examples are: pēṭ-ōṭ 'stomach and other organs' (from pēṭ 'stomach'), ghar-or 'house and household effects' (from ghar 'house'), chūrī-ūrī 'knives and things, cutlery' (from chūrī 'knife'), khīrā-ūrā 'cucumbers and other vegetables' (from khīrā 'cucumber'), ām-ōm 'mangoes etc' (from ām 'mango'). The rule for deriving the echo can be economically stated as follows:

$$\#\# (C)\ V\ X \longrightarrow \emptyset \begin{matrix} V \\ [+\ back] \end{matrix} X$$

The rule states that initial consonants are dropped, a front vowel is changed to a back vowel, with other features (length, nasality, and height) unchanged, and the rest of the word repeated. Back-vowels, as the symbolism suggests, are unaffected by the rule. Further details concerning the social use of the construction, its applicability to other word-categories, and its analogues in other Indian languages can be found in K.M. Tiwary (1968).

A related construction in Bhoj is one which duplicates a masculine noun in -ā, but changes the masculine ending into the feminine -ī: eg. laṭhā-laṭhī 'fighting with sticks' (where laṭhā = 'stick'), jutā-jutī 'shoe-beating' (jutā = 'shoe'), and others, mostly restricted to the semantic field of 'fighting'.

c) Reduplication and Onomatopoeia: Another form of word-play which deserves brief mention here is the predilection for sound-symbolism, especially among verbs and adverbs, most of which are of non-Sanskritic

origin. These can be grouped as follows:

i) Onomatopoeia Proper, including verbs like phũk 'to blow', hāk 'to cry out', chīk 'to sneeze', hicuk 'to belch', and adverbs, usually in reduplicated form, like jham-jham 'profusely' (as in rain falling profusely), han-han 'swiftly' (as with the flight of a bird, or movement of a train), ghaṭar-ghaṭar 'a manner of drinking with speed and relish' etc.

ii) Quasi-onomatopoeic reduplication: While not being strictly onomatopoiec, the following do have some similarities with above group: reduplicated verbs like phac-phacā 'to be drenched', kac-kacā 'to be startled', dhuk-dhukā 'to rise and fall'; partially reduplicated verbs like cul-bulā 'to be eager to move away', kas-masā 'to be ill', and har-barā 'to be afraid to answer'; and partially reduplicated nouns like phaṭ-phuṭ 'cracking and splitting', kaṭ-chaṭ 'cutting and trimming', hāk-dāk 'shouting and yelling'. This last set differs from echo-compounds in that both elements are free forms, and are roughly equivalent in meaning.

2.3 Bhojpuri as an (S)OV language: Bhoj is an SOV, postposing language, with the verb consistently in final position in unmarked sentences:

1. ham phal tūrab.
   I fruit break.1sg.fut
   'I will pluck fruit.'

There are rare instances in which the basic order is violated - marked sentences, usually having an intransitive main verb, in which the topicalised elements are postponed till the end of the sentence:

2. gail lōg.
   go.3pl.past people
   'The people went.' (focussing on the people)

Bhoj has word order patterns which are almost identical to those of other members of the Indic family, the chief ones being outlined below.

2.3.1 Noun Modifiers In Bhoj these consistently precede the head noun:

a) Adjectives precede nouns:

3. ū ēgō sunnar meharāru dekhalas.
he one beautiful woman see.3sg.past
'He saw a beautiful woman.'

In sentences in which the adjective is used as predicate to a copular verb, the (subject) noun precedes it, and the sentence order remains S Comp V:

4. meharāru sunnar bāṭī.
woman beautiful be.3sg.pres
'The woman is beautiful.'

b) Genitives also precede the head noun. In 5 below the head of the first NP paharadar, and of the second NP jiu are each preceded by a genitive:

5. rājā ke paharādār ōkar jiu bacaile.
king GEN guard(n.) he.GEN life save.3sg.past
'The king's guard saved his life.'

c) Relative Clauses typically precede the main clause containing the NP of which the relative clause is an expansion:

6. je ban mē raure sikār karile, te ban mē
which forest LOC you(+R) hunt do.2sg.HAB that.CORR forest LOC
rahile.
live.1sg.HAB
'I live in that forest in which your honour hunts.'

d) For titles, kinship terms, and honorific particles, the proper noun comes first: eg Singh-jī 'Mr Singh' (where -jī expresses respect), Rām-Rāj 'King Ram' (literally 'Ram-king'), dharti māī 'Mother Earth' (literally 'Earth-Mother'), Gōpī māmā 'Uncle Gopi' (literally 'Gopi-uncle') etc. This might seem to be a violation of the rule that qualifiers precede head-nouns, but there seems to be a good case for considering the title to be semantically more important than the personal name, as suggested

by Ganhi Mahatma and Victoria Queen, which are un-learned Bhoj forms for Mahatma Gandhi and Queen Victoria respectively. [3]

2.3.2 Verb Modifiers: In Bhoj these usually follow the main verb, though there are a few exceptions.

a) Auxiliary verbs follow the main verb:

7. ū jae pāwal.
   she go.INF able.3sg.past
   'She was able to go.'

b) Interrogative verb modifiers in yes/no questions take the form of rising intonation on the final syllable of the sentence, or the question word ka 'what' after the verb:

8. ū kukur dekhalas ka?
   he dog see.3sg.past INTERROG
   'Did he see a dog?'

There is, however, a less colloquial variation in some dialects, in which the interrogative particle occurs in sentence initial position.

c) Reflexives are expressed in two ways: the reflexive pronoun āpan 'self' may be used for all persons in both the singular and plural, and always occurs before the noun it modifies (in the sense of 'one's own X'); while auxiliary verbs lē- 'to take' and dē- 'to give' are used to indicate whether the action expressed in a sentence is for the benefit of the agent or not.

9. ū āpan kām kailas.
   he REFLEX work(n.) do.3sg.past
   'He did his own work.'

10. ū kām kar delas.
    he work(n.) do 'give'3sg.past
    'He did the work ' (not necessarily for his own benefit).

11. ū kām kar lelas.
he work(n.) do 'take'3sg.past

'He did the work' (which was to his benefit).

The use of these'local' auxiliaries is discussed further in 2.5.2.

d) Negatives: Negative particles modifying the verb, however, always precede the main verb:

12. ū laikī nahi paṛh sakelā.
that girl not read able.3sg.HAB

'That girl cannot read.'

2.3.3 Other typological features: a) Adpositions: Bhoj employs postpositions rather than prepositions:

13. bāgh jangal se nikral.
tiger jungle ABL emerge.3sg.past

'The tiger emerged from the jungle.'

There is only one exception, binā 'without', which may occur as either a postposition (eg Rām binā 'without Ram'), or as a preposition (eg binā bolawalē 'without invitation').

b) Comparison of adjectives: The standard of comparison always precedes the comparative adjective. In 13 laikī se is the standard of comparison, and choṭa the comparative adjective:

14. laikā laikī se choṭa bāṭī.
boy girl ABL short be.3sg.pres

'The boy is shorter than the girl.'

c) Use of prefixes: These are scarce in Bhoj - Tiwari lists 68 suffixes (denoting various categories like 'agent', 'causative', 'diminutive', 'feminine' etc), as against 17 prefixes, most of which are Perso-Arabic loans (eg har 'every, each', bē 'without', khus 'happy') which have become opaque and lexicalised.

d) Co-ordination: Conjoining of two or more nouns or adjectives is effected by the use of a particle aur or au, which occurs before the last of the conjoined elements, eg. Mōhan aur Sōhan 'Mohan and Sohan' (these are proper names). Aur 'and' and baki 'but' may also be used as sentence co-ordinators in much the same way:

15. pānī barsal baki ghās hariar nahī bhail.
water pour.3sg.past but grass green not become.3sg.past
'It rained, but the grass did not become green.'

Gapping (the deletion of an identical verb in conjoined sentences, as in I went, and John too) is rare in Bhoj, but if it does occur, the first verb is usually deleted:

16. ham aur hamār bahini ainī.
I and I.GEN sister come.1pl.past
'My sister and I came'

However, the more usual means of expression would be to retain the verb ainī in the first clause as well: ham aini, aur hamār bahini ainī.

On account of these patterns, Bhoj can be said to have many of the features one expects from an OV language (after Greenberg (1966) and Lehmann (1973)), though it does have quite a few atypical features. Its preposed negatives, and co-ordination patterns are more commonly associated with VO languages, while its reflexives and interrogative verb modifiers have some features characteristic of OV languages, and others characteristically VO.

We now proceed to describe the major morpho-syntactic structures of Bhoj. Although not all of these are of equal importance later on in the text when language change and obsolescence are under discussion, it is nevertheless useful to have an overview of the Bhojpuri of India, before proceeding to examine the formation of a South African variety of the language. Where a syntactic feature or morphological pattern is focussed on again, it will be either cross-referenced back to this chapter, or briefly summarised as a prelude to further discussion.

2.4 The Noun Phrase: The basic (surface) order of elements is given by the formula:

$$NP \longrightarrow \left\{ \begin{matrix} (Det)(Adj)\ N \\ Pn \end{matrix} \right\}$$

2.4.1 Noun Forms: Bhoj nouns need not be classified into distinct declension groups, since the endings employed are the same for all nouns, irrespective of gender and stem form. A noun stem may end in a vowel, usually /ɑ/, /i:/, /i/, and /u/, and less commonly /e:/,/o:/, /u:/, /ɑ̃:/, and /i:/, or in a consonant. Some examples are:

rastā	'road'	ãkhi	'eye'	bhalū	'bear'	pānī	'water'
dāg	'stain'	bār	'hair'	hāth	'hand'	bagh	'tiger'

There are usually two stem forms - a short form (as given above), and a long form, marked by a semi-vowel as being 'familiar', 'emphatic', 'definite', or 'contemptuous'. Which one of these categories is intended can be deciphered by both context and tone. A few examples are:

rāt	'night'	ratwā	'the night' (usually 'definite' or 'emphatic')
māī	'mother'	maiyā	'the mother' (usually 'familiar')
pēṭ	'stomach'	peṭwā	'the stomach' (any of the above categories)

The alternation y / w is phonologically conditioned, [j] occurring after front vowels, and [w] elsewhere (ie. after consonants or back vowels), with a few exceptions.

A third form which Grierson and Tiwari call the 'redundant' form, exists in some dialects, and for some nouns only, in which [wɑ:] is added to the long form of the noun, with emphatic effect: thus, nāu 'barber, nauwā (long form), nauwawā (redundant form); māli 'gardener', maliyā (long form) form), and maliyawā (redundant form).

In addition to the primary noun stems, there are a number of derived nouns:

With -i: khēli 'game' (from khēl- 'to play')

With -rī: pujārī 'worshipper' (from pūjā 'prayer')
With -ai: dekhai 'act of seeing' (from dēkh- 'to see')
With -nī: caṭnī 'a relish' (from cāṭ- 'to lick')

2.4.2 Gender: Gender distinctions are not as regular as in Standard Hindi where any noun is assigned to a grammatical gender (male or female), which determines the form of the adjective or the postpositional case-marker ka. In Bhoj agreement between adjective and noun is not compulsory, nor is the case marker ke inflected for gender (for examples see 2.4.5). Certain gender patterns among nouns can, nevertheless, be discerned:

a) Masculine in -ā: The -ā ending usually signifies masculine gender, a category which includes masculine animate beings, as well as inanimate objects, having the following sub-divisions:

i) Referring to male animate beings, and having feminine equivalent in -ī: eg bēṭā 'son', bēṭī 'daughter'; laikā 'boy', laikī 'girl'; murgā 'cock', murgī 'hen'; ghōṛā 'horse', ghōṛī 'mare' etc.

ii) Referring to inanimate objects, with diminutive equivalent in -ī: eg ḍōlā 'palanquin', ḍōlī 'small palanquin'; ghaṇṭā 'bell', ghaṇṭī 'small bell', mircā 'chilli', mircī 'small chilli' etc.

iii) Referring to inanimate objects, and having no feminine grammatical equivalent: eg tablā 'small drum' (musical), loṭā (type of drinking vessel), dānā 'grain', rastā 'road' etc.

b) Masculine in -ī:

i) Referring to masculine animate beings, with feminine equivalent in -īn: eg. telī 'oil-presser', telīn 'wife of oil-presser'; dhobī 'washerman' dhobīn 'washer-woman' or 'wife of washer-man'; hāthī 'elephant', hāthīn 'female elephant' etc.

ii) Referring to masculine animate beings, inanimate objects, or abstract entities, and having no feminine grammatical equivalent. These, I class

as masculine because in the few dialects which have concord of adjective and noun, the adjective form here is the masculine. Eg lakiṛī stick', sādī 'wedding', admī 'man', khetī 'farming', gārmī 'warmth' etc.

c) Unmarked Masculine:

i) With feminine equivalent in -īn(i): eg sonār 'goldsmith', sonārīn 'wife of a goldsmith'; nāg 'serpent', naginī 'female serpent', bēsaram 'shameless man', besarmīn 'shameless woman' etc.

ii) With no feminine equivalent: These I classify as masculine because only the masculine form of the adjective may qualify them. Eg dhan 'wealth', marad 'man', bajār 'market', dīn 'day' etc.

d) Feminine Nouns: I classify as feminine those nouns which may govern the feminine form of the adjective, even though in most dialects of Bhoj (excluding eastern dialects) the masculine (unmarked) form of the adjective is more often used. Thus laikī 'girl' is feminine because baṛī laikī 'a big girl' is possible, even though baṛā laikī is the more common form in most varieties. By the same token laikā 'boy' is masculine because *baṛī laikā (with the feminine form of the adjective) is impossible. In addition to the feminine classes corresponding to the masculine classes outlined above, there is an unmarked group (which may govern feminine adjectival forms) eg. aurat 'woman', sās 'mother-in-law', patōh 'daughter-in-law' etc.

2.4.3 Number: Plurality for nouns is denoted by means of a suffix -n(h) or -an(h) for many nouns, or by periphrasis, with the addition of the free form lōg 'people' (for human or human-like nouns) or sabh 'all' (for inanimate objects or for pronouns). Thus baccā 'baby' has the plural forms baccan , baccā lōg, and sometimes even baccā when the context makes it clear that the noun is plural, especially if it is preceded by a numerical qualifier. Doubly plural forms, with both an -n(h) ending plus lōg occasionally occur.

2.4.4 Case: Semantic roles of 'agent', 'patient', 'beneficiary', 'instrument' etc

are signified by a system of postpositions, rather than by case marking. There are a few relic case endings in -ẽ, -an, or -anhi for the instrumental, restricted to a few words: dãtẽ or dātan 'with teeth', bhūkhẽ or bhūkhan 'on account of hunger'. Locative endings in -ē or -ẽ are more widespread: gharẽ 'at home', bajarē 'in the market', etc.

Most often there is a distinction between the unmarked nominative, and the other cases (the 'oblique' cases) which take an optional -ē ending, plus a postposition. These postpositions are the chief carriers of role distinctions, though the difference between agent and patient is sometimes expressed by word-order alone. Postpositions differ from case endings in that they are not suffixed to the nouns they qualify. This can be demonstrated by invoking the rule of the short antepenult (see 2.1.3). Whereas the addition of suffixes, including case-endings, results in the shortening of long root-vowels in the antepenultimate syllable (eg pānī 'water', paniyā 'water' (definite), paniyē 'water' (definite, oblique)), this does not happen when a noun is followed by a postposition (eg pānī mē, not *pani mē).

The main postpositions are:

a) ke - the oblique marker par excellence. It occurs with [+ human] patients (to be designated 'ACC'), with nouns whose semantic role is that of 'experiencer', 'recipient' or 'beneficiary' (all designated 'DAT'), as well as 'possessor' (or 'GEN'). Although nouns do not have separate dative and genitive postpositions,[4] a distinction needs to be made between the two cases on account of pronominal paradigms, which do treat them differently. The dative postposition remains ke, but for the genitive kar (or some variant) is used instead - eg. ō-ke 'to him, for him' etc, versus ō-kar 'his'.

b) se - covering the roles 'source', 'comitative', 'instrument' and 'force' (to be designated 'ABL').

c) mẽ - covering 'inessive', and 'temporal' (and designated 'LOC').

d) par - covers 'adessive' (and designated 'AD').

There are no real surprises here, and these postpositions correspond fairly closely to the Indo-European style of inflections, maintaining a 'dative of possession', and having the same marker for ablatives, instrumentals, and comitatives. There is some evidence that postpositions might be on the way to being treated as suffixes. This is particularly true of pronoun plus postposition sequences, which are intuitively thought of as one unit eg. hamke 'to me', ōse 'from him', ēmẽ 'in this' etc. In addition there is the unusual sequencing of the postpositions mẽ (locative) and se (ablative) as follows: ākhī mẽ se 'from within the eye' or just 'from the eye', ōmẽ se 'from inside that', etc. As these postpositions can be found in many of the example sentences in the rest of the chapter, we limit ourselves to a few illustrations here.

17. ham Rām ke dekhlī̃.
I Ram ACC see.1sg.past
'I saw Ram.'

18. baccan ām turle
child.PL mango cut.3pl.past
' The children plucked mangoes.'

19. rāt mẽ garjat rahal.
night LOC thunder.PP be.3sg.past
'It was thundering in the night.'

In 17 the human object requires the postposition ke, while 18 shows the non-use of ke with non-human objects. There are a few other postpositions like talak 'up to', kane 'near', which are not as basic as those set out above. In addition there are many postpositional adverbs (ke) picche 'behind', (ke) lage 'near', (ke) kinārē 'beside', (ke) āgẽ 'in front of', (ke) khatin 'for the sake of' etc., which occur after the postpositions.

2.4.5 Adjectives: Adjectives resemble nouns formally in that most masc-

uline forms have -ā or are unmarked, and have feminine equivalents in -ī. Pairs like patrā (m) patrī (f) 'thin'; ujar (m), ujarī (f) 'white'; golā (m), golī (f) 'reddish'; lambā (m) lambī (f) 'long' occur. As mentioned previously, not all dialects utilise such gender distinctions systematically: in most dialects a phrase like 'a thin man' would be rendered patrā admī, with a masculine adjectival ending, while 'a thin woman' would be either patrī āurat or patrā āurat.

There are some invariant adjectives in -ī, which are used with both masculine and feminine nouns, eg alsī nau 'lazy barber' (m), alsī laikī 'lazy girl' (f).

Adjectives do not generally inflect for case and number, though some Eastern dialects, under the influence of Std Hn, use -e as both plural and oblique adjectival form. Like the noun, most adjectives admit a long form in -kā (paralleling that of -wā for nouns), and in some dialects, a 'redundant' form in -kawā. The long forms carry the same connotations as the noun forms ('definite', 'emphatic', 'familiar', 'contemptuous'), and may be used in conjunction of the noun, though the noun is usually left unmarked if the adjective is already in long form. Thus chotkā baccā 'the small child' is both 'definite' and 'emphatic'; but chotkā baccawā (with both adj and noun in the long form) and chotkawā baccā (with the adj. in the 'redundant' (or extra-long) form) are more emphatic.

For a brief discussion of the comparative form of adjectives see 2.3.3. The superlative form, a stylistic variant of the comparative, adds the phrase sabh se 'of all of' :

20. ũ laikā sabh se nīk hawe.
that boy all ABL good be.3sg.pres
'That is the best boy (of all).'

An adjective may be intensified by addition of thoṛā 'few, a little', baṛā 'big, great', or bahut 'much' or by reduplication:

21. kām baṛā acchā hẽ.
work(n) 'big' nice be.3sg.pres
'The work is very good.'

22. kaccā kaccā phal nice giral hẽ.
green green fruit down fall.PAST P be.3sg.pres
'Very green fruit have fallen to the ground.'

Where a string of adjectives occur the ordering: (Quantifier - Attributive - Size - Colour) is the most common one; eg. ẽk acchā baṛā lāl mōṭar 'a wonderful, big, red car'.

2.4.6 Pronouns: As full paradigms of Indian Bhoj pronouns, and patterns of simplification in South Africa are discussed in the following chapter we need not go into details here. The first person pronoun is ham, with plural forms hamahan/hamanī (or ham lōg, in some dialects, showing the same plural by periphrasis as nouns). The second person pronouns are tū (or tũ in some dialects) which is a (-R) or 'familiar' form, te (or tẽ) which is unmarked, while in western dialects the (+R) form rauwã is used. The third person forms are ī (proximal) and ū (distal), which are the same as the demonstratives.

Interrogatives, relatives and correlatives can be grouped together on account of their formal similarities, with the alternation /k/, /j/, /t/ in the initial consonant marking off a pronoun as belonging to one of these groups. Interrogatives begin with /k/ eg kaun 'which', kẽ 'who', kẽkar 'whose'; relatives begin with /j/ eg jaun 'that which', jẽ 'he who', jẽkar 'he whose'; and correlatives with /t/ eg taun 'which', tẽ 'who', tẽkar 'he whose'.

23. Jab hamarā hāl ke sunabe ta(b) ãkhi mẽ se lōr
when I. GEN news DAT listen.2sg.fut then eye LOC ABL tears
girāwe lagabe.
fall.CAUS.INF. begin.2sg.fut
'When you listen to my news, then you will begin to drop tears from your eyes.'

Sentence 23 shows the use of /j/ and /t/ forms as relative and correlat-

ives. The relative clause is the subject under discussion in 2.6.2.

Bhoj has one reflexive pronoun, āpan (oblique apane) denoting 'self', or 'one's own', which may be used with all persons.

24. ham āpan pōthī lēb.
I REFLEX book take.1sg.pres
'I will take my book.'

To replace the reflexive here with the ordinary genitive form of the first person pronoun (hamār 'my') would be ungrammatical, since the nominative form of a noun or pronoun may not co-occur with an oblique pronoun with the same referent, in the same clause. Whereas āpan usually has a genitival/reflexive function, its oblique form apanē denotes 'by oneself', 'alone':

25. ham pōthī apanē lēb.
I book REFLEX.OBL take.1sg.pres
'I will take the book myself.'

2.4.7 Determiners: The hierarchical order for determiners is given by the formula (with the proviso that the definite particle -wā is attached to nouns).

$$\text{DET} \longrightarrow (\text{Limiter}) \left(\left\{\begin{matrix}\text{Demonstratives}\\ \text{Possessives}\\ \text{Definites/Indefinites}\end{matrix}\right\}\right) \left(\left\{\begin{matrix}\text{Totaliser}\\ \text{Numeral}\end{matrix}\right\}\right)$$

The demonstratives are sub-divided into a proximal set, with nominative forms ī (sg), ihãka (pl), and oblique ē (sg), inhani (pl); and a distal set , with nominative forms ū (sg), huãka/uhãka (pl), and oblique ō (sg), unhan(i) (pl). These serve as deictic pronouns, and as demonstrative adjectives: eg Ī tō lakiṛī hē 'This is a stick' or 'It is a stick' (deictic) and ū admī 'that man' (adjectival).

The 'totalisers' are sab(h) 'all' and kulhi 'all, whole', the indefinites kaono 'some, any', ēk 'one, a', the limiters khalli/kewal 'only'. The numerals are based on a decimal system of counting, and in addition to the cardinals, consist of a host of ordinals, multiplicatives, collectives, fractionals, distributives, subtractives, proportionals etc. (see Tiwari 1960:116-125).

2.4.8 The Classifiers: These are bound morphemes thō/thē (in Western Bhoj and Awadhi), and gō (in Eastern Bhoj), used after numerals to denote discrete entities. The sequence numeral + classifier (+ noun) signifies 'x units of Y', eg. tīn thō ādmi aile means 'three men came (individually)' whereas absence of the classifier would suggest that they came as a group. A frequently used formula in stories Ēk-gō ādmi rahal 'There was a (certain) man', where the classifier serves to isolate the hero from other persons. As a final example the absence of a classifier in the sentence Rānī ke pā̃c caukidār hamke pakar lihalaī̃ 'The queen's five watchmen took hold of me' suggests that the five men acted as a collective unit, and not as individuals. Classifiers may occur with definite or indefinite numerals (an example of the latter being bisani gō 'scores of'), and with some interrogatives (eg. ketna/kai thō 'how many'), but not with cardinals or collective nouns.

2.5 The Verb Phrase:

2.5.1 Verb Types: In Bhoj there is no distinction between roots and stems in the classical sense of roots being abstract forms underlying all verb stems, which are derived from roots in a systematic manner. Instead one finds a simple system of basic verb forms to which endings may be attached directly, and which may themselves occur as free forms. Formally there is nothing that distinguishes intransitive base verbs from transitives. Thus kar- 'to do', rakh- 'to keep', lē- 'to take', dēkh- 'to see', pahir 'to wear', pukār- 'to call' (all transitive), are indistinguishable from the intransitive forms kā̃p 'to tremble', rō 'to weep', jā- 'to go', cal- 'to move'. It is only in the endings of the past tense that the classes differ formally.

As the examples suggest, most verb stems (as I shall call them) are monosyllables. There are two important categories of derived verbs: a) the 'first causatives' which are formed by the addition of -ā(w) to the basic forms, whose function is to make transitive stems intransitive, or to turn transitive stems into causatives. Thus the first causative equivalent of gir 'to fall' (intrans) is girā- or girāw (depending on

regional dialect) whose meaning is 'to drop' (trans), and of dēkh- 'to see' (trans) is dekhā(w)- 'to show' (trans).

b) the 'second causatives' which are formed by the addition of -wā(w) to the basic stem, with the (transitive) meaning of 'cause/make X do Y'. The second causative forms for the above verbs are girwāw- 'cause X to drop Y' (or 'cause X to cause Y to fall'), and dekhwā(w)- 'cause X to show Y' (or 'cause X to cause Y to see').

26. mistrī hamanī khatir kursī ban-ā-i.
artisan we.OBL for chair make.intrans-CAUS-3sg.fut
'The artisan will make a chair for us.'

27. Hamār bēṭā mōṭar ghum-āw delas.
I.GEN son car move.intrans-1ST CAUS 'give'3sg.past
'My son moved the car.' (Literally 'My son caused the car to move.')

28. tū āpan bēṭā se mōṭar ghum-wāw dehaliu.
you REFLEX son ABL car move.intrans-2ND CAUS 'give'2sg.fem.past
'You (f) made your son move the car.' (literally 'You caused your son to cause the car to move.')

2.5.2 Tense and Aspect: There are three distinct tenses - present, past, and future, with aspectual distinctions of 'progressive' and 'perfective'. Since these have been well maintained in South Africa, and since the paradigms show tremendous social and regional variants, we will list only the past tense forms here, The data below represents the forms current in Eastern Bhoj, taken from Tiwari (1960:172) for the verb dēkh- 'to see'

	Sg	Pl
1.	dekhalī̃	dekhalī̃-ja
2.	dekhalē̃ (-R) dekhalā́ dekhalī̃ (+R)	dekhalasa(n) (-R) dekhalā́ dekhalī̃ (+R)
3.	dekhalasi (-R) dekhalạni/dekhale dekhalī̃ (+R)	dekhalasa(n) (-R) dekhal dekhalī̃ (+R)

All Bhoj paradigms have feminine forms as well, most often in the 2nd and 3rd persons. Feminine forms for the first person indicate a female speaker, for the second person, a female addressee, and for the third person, a female referent different from both addresser and addressee. In the past tense the feminine forms are the same as for the masculine forms given above, with the following exceptions: the 2nd person unmarked form (sg and pl) is dekhalu, not dekhalá, the 2nd person pl (-R) form is dekhalusa(n), not dekhalasa(n), the 3rd person unmarked form is dekhalī in the sg, not dekhalani, and the 3rd (-R) pl form is dekhalisa(n), not dekhalesa(n).

The number of endings in all the main varieties of Bhoj (Western, North Eastern, South Eastern, Nagpuria), and all their sub-varieties are too many, and the differences too minute to be listed here. Not only does the final vowel differ from region to region, but the -l characteristic of the past is in some dialects replaced by [n] or [u].

The characteristic ending for the future is /b/ plus a vowel, though most dialects have an /i/ or /ihe/ in the third person. The present tense (usually non-progressive, and habitual) is characterised by retention of long stem-vowels, and the ending /l/ plus vowel.

Progressive and perfective aspect are denoted by use of the participle plus auxiliary. The present participle is formed by the addition of /at/ to the verb stem, and the past participle by /al/. The auxiliary verb that occurs is identical to the copula, with suppletive forms -bā (or haī in some dialects) for the present, rah- for the past, with either form possible in the future. Thus ham bānī 'I am' (or in more eastern dialects ham haī), ham rahab 'I will be' or, less commonly ham hob (with subjunctive ham hoi), and ham rahalī 'I was'. Some examples illustrating aspectual combinations follow:

29. ū rastā par dhīre cal-at bā.
he road AD slowly walk-PP be.3sg.pres.
'He is walking slowly on the road.'

30. Ham ghare jā-t rahab.
I house.OBL go-PP be.1sg.fut
'I will be going home.'

31. tū khub sut-al bāre.
you well sleep-PAST P be.2sg.pres
'You have slept well.'

32. Ēgō gāī rakh-ale rahale.
One.CLASS cow keep-PAST P be.3sg.past
'He had kept a cow.' (or 'He used to keep a cow.')

c) Duplication of Present Participles: This is a common pattern in the modern Indic languages, a kind of non-paradigmatic aspect denoting intensity, or frequency, of action. It has analogues elsewhere in Bhoj grammar, especially in the repetition of adjectives and adverbs for stylistic effect.

33. laikā bhāg-at bhāg-at aile.
boy run-PP run-PP come.3sg.past
'A boy came running (at speed).' (or 'from afar')

2.5.3 Counter-Factuals and Subjunctives: Counter-factuals usually refer to suppositions regarding past events, and are expressed by the addition of the suffix -it to the verb base, followed by the personal endings for the ordinary past (as given in 2.5.2). The simple past and counter-factual paradigms are thus identical in all respects (including feminine forms and honorifics), except for the replacement of past tense /l/ by /t/ for the counter-factual.

33. ja ū dēkh-it, ta ūhū rō-it.
if he see-3sg.CF then he.INCL cry-3sg.CF
'If he had sen it, he too would have cried.'

In sentences having aspectual forms, the counter-factual ending is attached to the auxiliary, while the main verb remains in participial form:

34. ja ham unhukā dēkhat rah-itī̃, ta māī ke kahale rah-itī̃.
if I he.OBL see.PP be-1sg.CF then mother DAT say.PAST P be.1sg.CF
'If I had been seeing him, I would have told mother.'

Subjunctives, on the other hand, express possibility, volition, requests, and doubt, all with a future connotation. The subjunctive paradigm does accordingly resemble, and overlap with, the future paradigm, formally and semantically (see the third person future in -ī). The subjunctive ending is ī (or ĩ) for all persons and numbers:

35. ōke bolāwa ki ham bāt kar-ī.
he.DAT call.CAUS.IMP that I speech do-SUBJ
'Call him, so that I can speak to him.'

36. ab ham thoṛa baiṭh-ĩ.
now I little sit SUBJ
'Now I will sit awhile.' (or 'Now I wish to sit awhile.')

2.5.4 Compound Verbs: In common with the other Indic languages Bhoj has a large number of compound verbs, made up of two verbs which may occur independently of each other as full verbs at other times. The first verb, expressed in stem form, carries the bulk of the meaning, while the second typically marks grammatical categories like 'inception', 'completion', 'chance', 'reflexiveness', 'intensity' or 'deixis'. Whereas the first could be almost any full verb, the second is drawn from a small set of verbs whose literal meanings have become grammaticalised in this construction. These 'operators' (as the second verbs are often referred to by grammarians) are: lē- 'to take', dē- 'to give', lag- 'to feel', ā- 'to come', jā- 'to go', ḍāl- 'to pour', 'to put in', par- 'to fall', cal- 'to move', and a few others which are less common.

37. sãpwā ke mār ḍallas.
snake.DEF ACC kill 'pour'3sg.past
'He killed the snake.' (or perhaps 'He butchered the snake.')

38. pōthī uhã rakh dēb.
book there place 'give'1sg.fut
'I will place the book there.'

39. tũ sōb pānī pī lelā.
you all water drink 'take'2sg.past
'You drank up all the water.'

In 37 the operator dal- conveys intensity of action, the full verb mārlas alone would be grammatical, though less emphatic. The 'local' auxiliaries in 38 and 39 signify completion of action, the first dē- indicates that the action is not necessarily for the benefit of the agent, while the second, lē-, which is more reflexive in function than the other operators, is used when the action expressed by the verb is to the agent's satisfaction or advantage. The semantics of these compound verbs, and restrictions on the operators is much too vast a topic to be dealt with here, and the interested reader is referred to Hook (1974).

2.5.5 Modals: Modal auxiliaries differ from the operators defined above in that although they have a clearly defined meaning, they do not occur on their own (unless the full verb is understood to be deleted under anaphora). The modals of Bhoj are: sak- 'can, to be able', pā - 'to manage, be able', cāh- 'to wish', while the inceptive lag- 'to begin' and the completive cuk- 'to finish' can be listed here on account of their formal similarities with the modals. These verbs take the same endings as other verbs, and are preceded by a full verb in either their stem or infinitive form. Unlike the 'operators', modals can be used quite idiomatically in negative sentences.

40. bahini kaprā nā dhō sak-elā
sister clothes not wash able-3sg.HAB
'My sister is unable to wash clothes.'

41. ham khā cuk-lī
I eat finish-1sg.past.
'I finished eating.'

42. ū abhi bōle cāhat bā.
he now speak.INF wish.PP be.3sg.pres
'He wants to speak now.'

We shall see in chapter 5, that whereas full verbs and operators remain fairly stable in South African Bhojpuri, the modals have been susceptible to many changes.

2.5.6 Verbal Noun and Infinitive: There is, unfortunately little agreement in the literature concerning terminology and descriptions of these categories in various dialects of Bhoj. Each of the following have been described as infinitives and/or verbal nouns by some writer or other:

a) verb stem + Ø : This form occurs in the conjunctive construction (see 2.6.3), and as an imperative (see 2.6.6). It seems best to characterise it as "stem" rather than infinitive or verbal noun.

b) stem + -e: This seems to me the only infinitive in Bhoj, functioning as complement of verbs such as jā- 'to go', ā- 'to come', dē- 'to give', cāh- 'to want' eg. ham sūte jāb 'I will go to sleep', where the infinitive form sūte precedes the main verb jāb 'I will go'. The infinitive is also used in purposive constructions, where it is always followed by the dative postposition ke (see 2.6.4).

c) stem + -ab: This is a verbal noun (with oblique form -be in many dialects) which occurs as the complement of the copula, or co-occurs with adjectives eg. sūtab nīman hōlā 'Sleeping is good/ It is good to sleep'. Tiwari (1960:194) notes that this ending was becoming obsolete in South Eastern dialects by 1960.

d) stem + -al: This verbal noun form is noted in all dialects by Grierson but seems to be more popular in some areas, where it is used instead of the stem + -ab form, eg parhal karalī̃ 'I kept on reading' (where the combination of kar 'to do' and verbal noun suggests a frequentative sense).

2.6 The Major Syntactic Constructions:

2.6.1 Passives: The passive construction is not as frequently used in Bhoj as in English, especially when the agent is known. It involves the switching of agent and patient (in terms of word order and case/postpositional marking), replacement of the main verb by its past participial form, and addition of the auxiliary verb (or 'operator') jā- 'to go',

whose endings show agreement with the new subject (ie. the patient). The agent is most often deleted.

43. dūdh mē bhei ke rōṭī khā-il jālā.
milk LOC soak CONJ bread eat-PAST P 'go'3sg.HAB
'Bread is eaten after having been soaked (literally 'after soaking') in milk.' (from Tiwari 1960:164)

44. Ōkar khēt hamār ghar se dēkh-al jālā.
he.GEN farm I.GEN house ABL see-PAST P 'go'3sg.HAB
'His farm can be seen from my house.'

As both these sentences may suggest, the passive is more idiomatic in such 'universal' and/or 'ability' statements lacking an agent, and is more frequently used therefore in the habitual present tense than any other. In the past tense it is more usual to use active sentences, containing verb compounds (most often having the 'operator' jā- 'to go' in its past form) which convey a perfective and passive sense.

45. sab kaṭa nikal gail.
all thorn come-out 'go'3pl. past
'All the thorns were removed' (literally 'came out').

46. jal se bhari gaile tāl-talāī
water ABL fill 'go'3pl.past lake-pond.
'The lakes and ponds are filled up with water.' (from Tiwari 1960:164)

In 45 and 46 nikal gail and bhari gaile are compound verbs which are formally active, but convey a perfective and passive meaning.

2.6.2 Relative Clauses: The usual construction is a correlative one, in which the relative clause (introduced by the relative pronoun je or jaun) precedes the main clause which itself is introduced by a correlative te or taun. The 'equivalent' NP in the main clause is usually deleted.

47. jaun laikā tē liāile, taun hamār ghaṛi corailas.
which boy you bring.2sg.past that.CORR I.GEN watch steal.3sg.past
'The boy you brought stole my watch.'

In some sentences the ordinary third person pronoun may replace the corr-

elative pronoun:

48. jaun bānar hamani āj dekhlī-jā, ū jangal mē cale
which monkey we.pl today see.1past-pl it jungle LOC move
gail.
'go'3sg.past
'The monkey which we saw today went away into the jungle.'

Sentence 49 exemplifies an ordinary (non-correlative) construction:

49. ū sādhu jē bāhar hē, ū kahā se ail
that holy-man who outside be.3sg.pres he where ABL come.PAST P
hē?
be.3sg.pres
'Where has that holy man who is outside come from?'

In some instances the relative clause is postposed. This (paratactic) order seems to apply to relatives used appositively, being added on almost as an after-thought:

50. ū manaī hathiyā dekhalani,....jaun pēṛ tūrat rahale.
that man elephant.DEF saw.3sg.past which tree break.PP be.3sg.past
'The man saw the elephant....which was tearing down the tree.'

2.6.3 Co-ordination: The particles au(r) and baki(r) ('and' and 'but' respectively) are used as sentence co-ordinators, occurring in-between the conjoined sentences:

51. hamār bhāī ailī, bakir nā rahalī.
I.GEN brother come.3sg.past but not stay.3sg.past
'My brother came, but did not stay.'

52. rājā kudle gelak aur dekhēlā ṭhīke bāt hē.
king run.PAST P 'go'3sg.past and see.3sg pres true word be.3sg.pres
'The king went running, and saw that this was true.'

When two verbs are sequentially related, it is more usual to use an absolute construction in which the first verb occurs in stem form followed by the particle ke (historically related to kar 'to do'), and the second in the usual finite form:

54. corwā admī lōg dēkh ke bhāg gaile.
thief.DEF man PT see CONJ run 'go'3sg.past
'The thief saw the men and ran.'

An alternate absolute form in some dialects is verb stem + -i. Thus, dēkh ke or dēkhi ke 'having seen'.

2.6.4 Complementation: Complements may be strung on paratactically to a main clause, with no special mark except for a brief pause:

55. ham dēkhīlā tũ pōthi paṛhālā.
I see.1sg.pres you book read.2sg.pres
'I see (that) you are reading a book.'

Alternately the particle ki may be used to introduce a complement, especially if it is in direct speech:

56. tab laikwā soclas ki ab ham baṛa admī bānī.
then boy.DEF think.3sg.past COMP now I big man be.3sg.pres
'Then the boy thought that he was now a "big shot".'

As 56 suggests the use of reported or direct speech occurs more frequently in Bhoj than in English. Complementation after verbs like sōc- 'to think', kaha- 'to say', paṛh- 'to read', and bōl- 'to relate' is expressed by direct speech, with verb forms, pronouns and other deictics unchanged. In complements expressive of compulsion or purpose, infinitive (stem + -e) clauses are most common.

57. nauwā rājā ke bār kāṭ-e gaile.
barber.DEF king GEN hair cut-INF go.3sg.past
'The barber went to cut the king's hair.'

58. ū pānī pī-ye cahatā.
he water drink-INF wish.3sg.pres
'He wishes to drink water.'

Compulsion constructions are characterised by infinitive + ke (dative postposition) + auxiliary verb:

59. ēk jane aur khōj-e ke pari.
one person more seek-INF DAT 'fall'3sg.fut.
'I will have to look for one more person.'

60. ā̄j sōb ghare jā̄-e ke bā̄i.
today all house.LOC go-INF DAT be.3sg.pres
'He has to go to all the houses today.'

Temporal clauses are introduced by jab and ta(b) ('when and 'then') in much the same way that correlative clauses are introduced by jaun and taun:

61. jab sanjhā̄ bhail, ta sādhu gāu se
when evening become.3sg.past then.CORR ascetic village ABL
aile.
come.3sg.past
'When evening fell, an ascetic arrived from the village.'

'If' clauses precede the result clause and are marked by ta 'then' in the second clause. In referring to past tense suppositions, the counterfactual endings in -t are used for the verbs in both clauses, as illustrated in 2.5.3. In suppositions concerning the future, the verb in the first clause or both verbs are in the subjunctive mood:

62. nōn bo-i ta kaisan ho-i?
salt plant-SUBJ then what be.SUBJ
'What will happen if I plant salt?'

As indicated in 2.5.3 the particle ja and correlative ta may be used to introduce the 'if' and the result clause respectively.

2.6.5 Interrogatives: Bhoj question particles usually begin with /k/, historically from the same source as English wh- words: kab 'when', kā 'what', kē 'who', kahā̃ 'where', kahe 'why', etc. Yes/No questions are usually marked by a rising intonation on the final syllable of the sentence, though an alternate means is to use kā 'what', also with rising intonation, in sentence-final position (or, in some dialects, initial position).

63. tohārā bahini ā̄j aibi?
your sister today come.3sg.fut.
'Will your sister come today?'

64. ī tohār bacca hawẽ, kā?
this your child be.3sg.pres what
'Is this your child?'

In information questions, the wh-word often occupies the same position as the word it might replace in an equivalent declarative sentence:

65. kē khirkī khōl delas?
who window open 'give'3sg.past
'Who opened the window?' (equivalent to Rām khirkī khōl delas).

66. tẽ kā kām karbe?
you what work(n) do.2sg.fut
'What work will you do?' (equivalent to ham sab kām karab 'I will do all the work').

2.6.6 Imperatives: Commands are effected by deletion of the second person pronoun (tū/tẽ/tohani) and use of one of the following verb forms: stem without an ending (for 'familiar' or 'contemptuous' usage), stem + -ā (for informal usage), and stem + -ĩ (for (+R) usage in some dialects). An example of the 'familiar' imperative, used to close acquaintances, or to signal anger or disrespect, follows:

67. bhīkhāri ke kuchū dē.
beggar DAT something give.IMP
'(Just) give the beggar something.'

2.6.7 Impersonal Constructions: Certain verbs, like lag-'to feel, to experience', mil- 'to get, to find', haī- 'to be' (only in possessive constructions), and par-'to fall' (as auxiliary) permit only dative subjects - ie. the 'deep' subject is always followed by the dative post-position ke - while the other noun governed by these verbs is unmarked.

68. ōke gussā lagal.
he.DAT anger feel.3sg.past.
'He was angry.' (literally 'to him-it felt-anger')

69. kisān ke auri paisā mīlī.
farmer DAT more money get.3sg.fut

'The farmer will receive more money.' (literally 'to the farmer-it will get-more money')

70. Mōhan ke baccan naikhī.
DAT child.PL be.neg.pres

'Mohan has no children' (literally 'to Mohan-it is not-children').

2.6.8 Subject Deletion: In 2.6.3 it was indicated that in both co-ordination and conjunctive constructions the subject occurs only once - ie. it can be considered to be deleted from one of the underlying clauses, usually the second. In addition there is another instance of subject deletion, where the subject is optionally omitted, provided context or verb-ending avoid any ambiguity of reference.

71. ā ke, dekhalas.
come CONJ see.3sg.past

'He/she/it came and had a look' (context making it clear who is being referred to).

72. Indar alas bāṭī. Kuchu nā karī.
lazy be.3sg.pres. anything not do.3sg.fut

'Indar is lazy. He will not do anything' (with the subject deleted in the second sentence).

NOTES

1. The transcription here, and in the rest of this work, is that of the IPA. The transliteration, however, follows traditions set by other works on Bhoj and other Indic languages written in English. It is hoped that the occasionally conflicting conventions are not confusing. The reader is referred to the key to the transcription and transliteration set out in the introduction.

2. Dvandva means 'two by two', tatpuruśa 'that man' (ie. illustrative of the type of compound it names), while bahuvrīhī 'much rice' refers to a third person who has much rice, ie. 'a rich man' (again illustrative of the compound it names).

3. However, some titles like paṇḍit 'priest', rājā 'king', and ṭhākur 'chief' often precede the proper name.

4. Shukla (1983) gives separate postpositions for North East Bhojpuri: kæ for the genitive, and ke for the dative.

CHAPTER 3

THE BHOJPURI LANGUAGE IN SOUTH AFRICA - A STUDY IN LANGUAGE COALESCENCE

In this chapter an attempt is made at delineating those features differentiating South African Bhojpuri from Indian Bhojpuri as outlined in chapter two, which are a result of contact between speakers of closely related dialects (eg. Western Bhojpuri, Eastern Bhojpuri, Magahi), or even closely related languages[1] (eg. Bhojpuri, Awadhi, Hindi). Subsequent changes resultant from contact with other South African languages as well as independent developments after separation from Indian Bhojpuri are discussed in chapter five.

3.1 The problem of names: We have already seen in chapter one that the official term "Hindi" used for the language in South Africa is quite problematic, since Hindi accounts for less than ten percent of the native speech of the first generation of North Indian emigrants. As a cover term for the various dialects in question "Hindi" might be as good as any other, but the chief drawback of this term is that it leads to unfair and unfavourable comparisons between the language spoken in South Africa and the Standard Hindi of officialdom in India (on such comparisons and negative attitudes to the language see chapter four). Villagers in India are quite vague about the name of their language, as emphasised by Grierson (1927:19) in the following anecdote about the difficulty of finding out the name of a village dialect or language:

> Just as M. Jourdain did not know that he had been speaking prose all his life, so the average Indian villager does not know that he has been speaking anything with a name attached to it. He can always put a name to the dialect spoken by somebody fifty miles off, but, - as for his own dialect, - "O, that has no name. It is simply correct language." It thus happens that most dialect names are not those given by the speakers, but those given by their neighbours, and are not always complimentary. For instance there is a

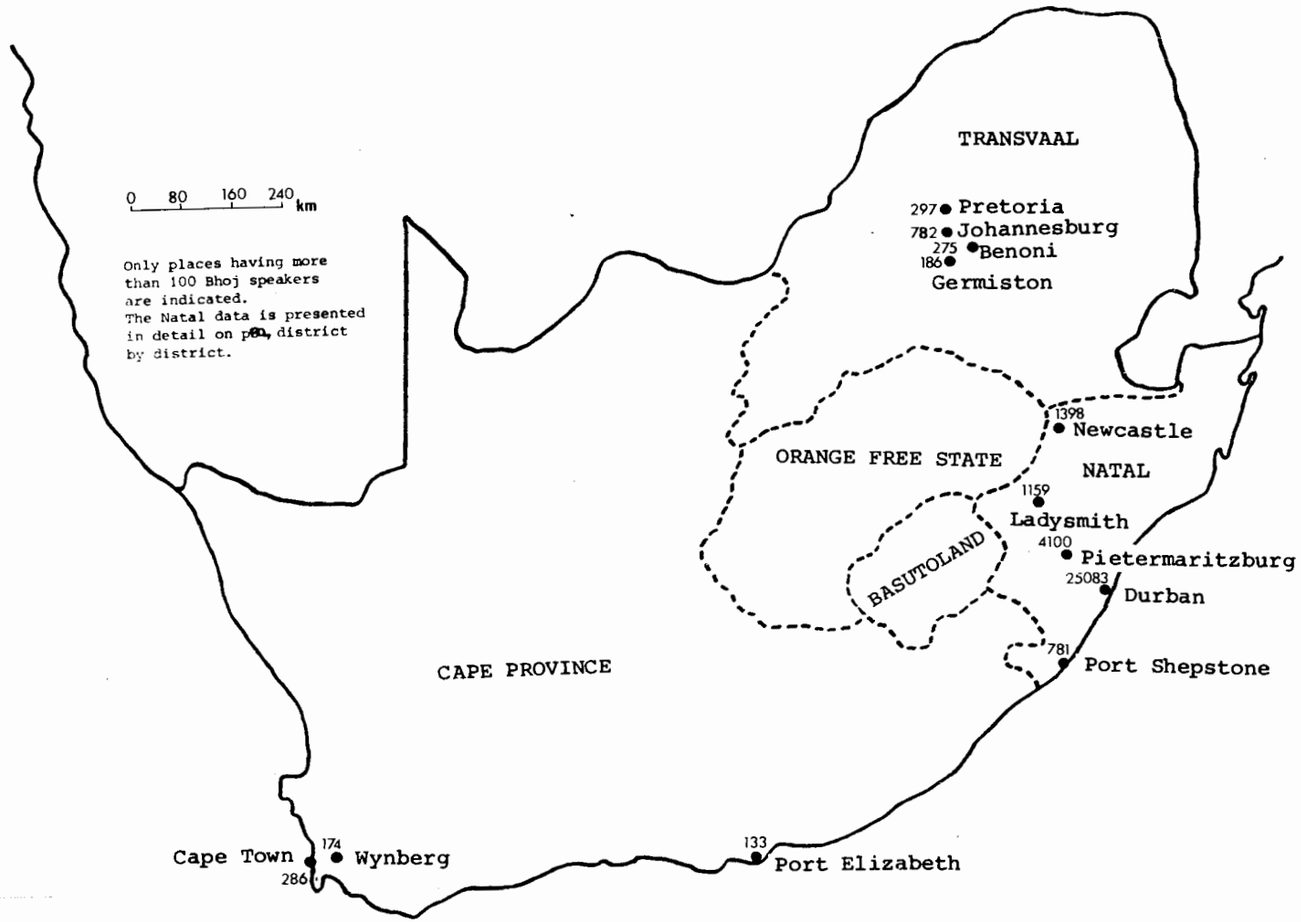
> well-known form of speech in the south of the Punjab called 'Jangali', from its being spoken in the 'Jungle' or unirrigated country bordering on Bikaner. But 'Jangali' also means 'boorish' and local inquiries failed to find a single person who admitted that he spoke that language. 'O yes, we know Jangali very well, - you will find it a little further on, - not here.' You go a little further on and get the same reply, and pursue your will-o'-wisp till he lands you in the Rajputana desert, where there is no one to speak any language at all.

In South Africa older speakers do use the terms Hindi or Hindustani for their language, but most often one hears the word Kalkatiā which could refer to either the people who boarded ship at Calcutta and their descendants, or the language that they speak. To be more explicit, the latter is sometimes referred to as Kalkatiā bāt (ie. 'language of the Kalkatiās). It is unfortunate that this commonly used epithet has not gained official recognition - it has the advantage of neutrality, giving preference to neither Bhojpuri, Awadhi, Hindi nor any other dialect, nor to any particular region, because although Calcutta was the port of embarkation, less than one percent of the original indentured labourers actually hailed from that city. In addition, the term carries connotations of historicity and the formation of a new identity as expatriates with a common future. The term Kalkatiā has its counterparts in the epithets Madrāsī (Madrājī, sometimes Mandrājī), referring to South Indians who had left via the port of Madras and to the Tamil language, and Baniyā, meaning 'merchant' in general, but in the local context referring mainly to Gujarati-speaking traders.

Not even the oldest speakers alive today use the names Bhojpuri or Awadhi. For the vast majority both names were unknown. The choice of label for the language in this study has, accordingly, not been easy. Linguists working with similar groups in ex-colonies elsewhere use the following labels: Mauritian Bhojpuri (Domingue 1971), Fiji Hindi (Moag 1977), Trinidad Bhojpuri (Ramesar 1978), and Guyanese Bhojpuri (Gambhir 1981). Because the earliest recruitment of workers (to Mauritius, the West Indian islands, and British Guyana) had taken place in Bihar and eastern Uttar Pradesh, the heartland of the

Bhojpuri language, the language spoken by these North Indians and their descendants in those new colonies is unambiguously Bhojpuri, albeit a mixture of related dialects. By 1860 recruiters, no longer finding it easy to attract enough people from Bihar, had to go further inland into western Uttar Pradesh, the domain of Awadhi and related Eastern Hindi dialects (see Saha 1970). The transplanted languages in South Africa and Fiji (to which recruitment began in 1860 and 1879 respectively), therefore show greater influence from languages other than Bhoj. Moag (1977:v), discussing the Fijian situation, claims that this influence makes the language quite unlike any particular regional Indian language or dialect, and accordingly prefers the term Fiji Hindi. I have settled on South African Bhojpuri (hereafter SABh), because, although the Awadhi influence is strong in some parts of the country, there is some justification for considering Bhojpuri to be the base, as I will show in the following sections. It would be incorrect, though, to think of SABh as an offshoot of Indian Bhoj alone.

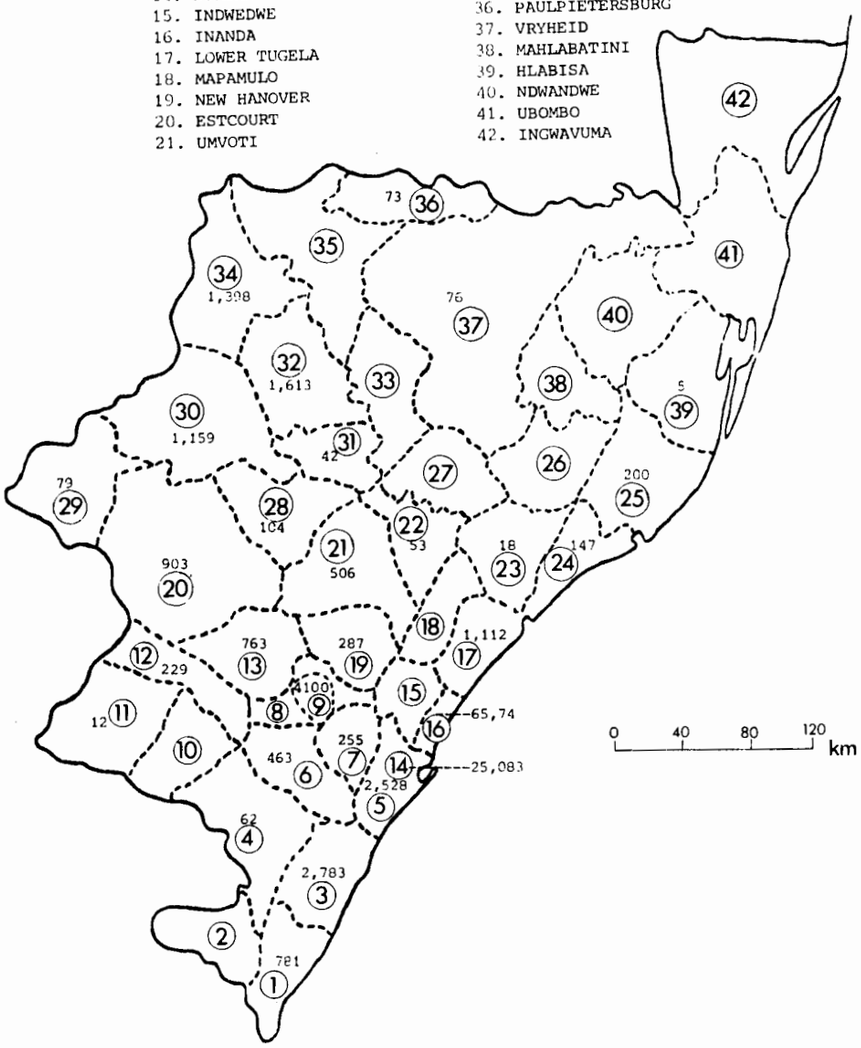
3.2 Location of South African Bhojpuri: The indentured Indian, initially restricted to working on sugar and sometimes coffee plantations in Natal found, on completion of his contract, that there was some scope for individual economic development. There were ready markets for small-scale farmers planting fruits and vegetables. Others resumed their traditional village occupations as tailors, cobblers, cooks, launderers and artisans. Some forms of employment which took them further into the interior were railway work (as early as 1866), and coal-mining in northern Natal. By the end of the nineteenth century there were as many as 1317 Indian workers on the mines, the earliest of whom had been recruited directly from India. Some left the British colony of Natal in search of employment in the gold and diamond fields of the Transvaal and Orange Free State, until the free movement of Indians was restricted at the turn of the century. By 1911 there were 3121 male and 1055 female Indian wage workers in Transvaal, a minority of whom were "Hindi" speaking, though this was a sufficient number for



Map 4 - Distribution of Bhojpuri speakers in South Africa - 1936.

Magisterial Districts

1. UMZINKULU
2. ALFRED
3. ALEXANDRA
4. IXOPO
5. UMLAZI
6. UPPER TUGELA
7. CAMPERDOWN
8. UMGENI
9. PIETERMARITZBURG
10. POLELA
11. UNDERBERG
12. IMPENDHLE
13. LIONS RIVER
14. DURBAN
15. INDWEDWE
16. INANDA
17. LOWER TUGELA
18. MAPAMULO
19. NEW HANOVER
20. ESTCOURT
21. UMVOTI
22. KRANSKOP
23. ISHOWE
24. MTUNZINI
25. LOWER UMFOLOSI
26. EMTONJANENI
27. UKANDHLA
28. WEENEN
29. UPPER TUGELA
30. KLIP RIVER
31. UMSINGA
32. DUNDEE
33. NQUTU
34. NEWCASTLE
35. UTRECHT
36. PAULPIETERSBURG
37. VRYHEID
38. MAHLABATINI
39. HLABISA
40. NDWANDWE
41. UBOMBO
42. INGWAVUMA



Map 5 - Distribution of Bhojpuri speakers in Natal - 1936.

Swami Bhawani Dayal to open a Hindi school in Germiston on a part time basis. Residence in the Orange Free State has been barred since 1891 and Indians living there in the twentieth century have been a mere handful. In the Cape the Indian population is still a small one. The earliest of these had been sent there directly from India, while others followed early (in the 1880's) to such jobs as catering and running retail stores after completing their indentureship in Natal.[2]

	1936	1951	1960	1970
Natal	56 489	86 588	116 566	114 584
Transvaal	2 228	2 091	7 053	1 730
Cape Province	1 533	460	2 458	171
Orange Free State	6	6	0	0
Total	60 276	9 145	126 067	116 485

Table 11 - Distribution of SABh speakers over a selected period.

Map 4 illustrates the distribution of Bhojpuri speakers for 1936. It shows that the Cape is negligible as a linguistic area for Bhojpuri, only two centres having more than a hundred speakers in 1936 - Cape Town and Port Elizabeth, and that only a few urban centres in the Transvaal need be considered. The 1970 figures reflect ethnic identity (descendants of Bhojpuri-speaking settlers), rather than linguistic reality, since they include a number of people, mainly children, who do not actively use the language. The increased figures for 1960 in the Cape and Transvaal are a result of the movements of younger people filling vacancies in both governmental

jobs - especially teaching - and private industry, who are often resident for short periods only (ie. less than a decade), and who very seldom use an Indian language in the new environment.

In the Transvaal and Cape, Bhojpuri has come into contact with both English and Afrikaans, and in rural areas with a Bantu language - Xhosa in the Cape, and to a small extent, with the Sotho languages (North Sotho, South Sotho, Tswana) in the Transvaal. In Natal it exists side by side with English, other Indian languages, Fanagalo, and Zulu. The language is on the whole equally distributed over both farm and city: in 1936 for example, there were 24,564 Bhojpuri speakers in rural areas, as compared to 35,712 urban-dwellers - a ratio of 5:7, the highest proportion of rural to urban speakers for a local Indian language. The other ratios were Tamil 3:5, Telugu 3:5, Gujarati 1:5, and Urdu 1:11. For 1970 the ratios were: Bhojpuri 1:3, Tamil 1:4, Telugu 1:5, and Gujarati 1:22, all clearly demonstrating a shift to urban areas.[3]

In most areas of Natal, Bhojpuri tends to be the second most-spoken language, after Tamil, whereas in the Transvaal and Cape it trails behind Gujarati and Tamil. Notable exceptions are in Northern Natal, KwaZulu, and a few rural areas in Southern Natal (eg. Louisiana in Port Shepstone). In the Northern Natal city of Newcastle, for example, the 1970 Census records 2391 "Hindi" (ie. Bhojpuri) speakers, 605 Gujarati, 415 Tamil, and 13 Telugu. Bhojpuri is, according to that Census, the most-spoken Indian language in the following areas: KwaZulu, Camperdown, Richmond, Pietermaritzburg, Dannhauser, Dundee, Glencoe, Newcastle, Ladysmith, Weenen, Lions River, Mooi River, and Estcourt. The figures for Pietermaritzburg, for example, are: 6,753 for "Hindi", 4,510 for Tamil, 1,368 for Gujarati, 324 for Telugu and 3,318 for other Indian languages in 1970.

3.3 Language Coalescence and the existence of dialects in Natal:

Minority languages like Rajasthani, Bengali, Nepali, and Oriya, whose total number was less than 6 percent of the original migrant population,