The Munda Languages

Edited by
Gregory D. S. Anderson

Routledge Language Family Series

THE MUNDA LANGUAGES

The Munda group of languages of the Austroasiatic family are spoken within central and eastern India by almost ten million people. To date, they are among the least well-known and least documented languages of the Indian subcontinent.

This unprecedented and original work draws together a distinguished group of international experts in the field of Munda language research and presents current assessments of a wide range of typological and comparative–historical issues, providing agendas for future research.

Representing the current state of Munda linguistics, this volume provides detailed descriptions of almost all of the languages in the family.

Gregory D.S. Anderson is Director of the Living Tongues Institute for Endangered Languages. His key publications include: *The Munda Verb: Typological Perspectives* (2007), *Auxiliary Verb Constructions* (2006) and *Language Contact in South Central Siberia* (2005).

ROUTLEDGE LANGUAGE FAMILY SERIES

Each volume in this series contains an in-depth account of the members of some of the world's most important language families. Written by experts in each language, these accessible accounts provide detailed linguistic analysis and description. The contents are carefully structured to cover the natural system of classification: phonology, morphology, syntax, lexis, semantics, dialectology, and sociolinguistics.

Every volume contains extensive bibliographies for each language, a detailed index and tables, and maps and examples from the languages to demonstrate the linguistic features being described. The consistent format allows comparative study, not only between the languages in each volume, but also across all the volumes in the series.

The Austronesian Languages of Asia and Madagascar
Edited by Nikolaus Himmelmann and Sander Adelaar

The Bantu Languages
Edited by Derek Nurse and Gérard Philippson

The Languages of the Caucasus
Edited by Alice Harris

The Celtic Languages
(new edition forthcoming)
Edited by Martin J. Ball and James Fife

The Dravidian Languages
Edited by Sanford B. Steever

The Germanic Languages
Edited by Ekkehard Konig and Johan van der Auwera

The Indo-Aryan Languages
Edited by George Cardona and Dhanesh K. Jain

The Indo-European Languages
Edited by Paolo Ramat and Anna Giacalone Ramat

The Iranian Languages
Edited by Gernot Windfuhr

The Khoesan Languages
Edited by Raïner Vossen

The Mongolic Languages
Edited by Juha Janhunen

The Oceanic Languages
Edited by John Lynch, Malcolm Ross and Terry Crowley

The Romance Languages
Edited by Martin Harris and Nigel Vincent

The Semitic Languages
Edited by Robert Hetzron

The Sino-Tibetan Languages
Edited by Graham Thurgood and Randy J. Lapolla

The Slavonic Languages
Edited by Bernard Comrie and Greville G. Corbett

The Tai-Kadai Languages
(forthcoming)
Edited by Anthony Diller

The Turkic Languages
Edited by Eva Csato and Lars Johanson

The Uralic Languages
Edited by Daniel Abondolo

The Munda Languages
Edited by Gregory D.S. Anderson

THE MUNDA LANGUAGES

Edited by
Gregory D.S. Anderson

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2008
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
711 Third Avenue, New York, NY 10017, USA

Routledge is an imprint of the Taylor & Francis Group,
an informa business

Typeset in Times New Roman by
Newgen Imaging Systems (P) Ltd, Chennai, India

British Library Cataloging in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
A catalog record for this book has been requested

ISBN10: 0–415–32890–X
ISBN13: 978–0–415–32890–6

To Norman

This page intentionally left blank

CONTENTS

This page intentionally left blank

ILLUSTRATIONS

MAPS

FIGURES

TABLES

BOXES

This page intentionally left blank

CONTRIBUTORS

Gregory D.S. Anderson is Director of the Living Tongues Institute for Endangered Languages (Salem, Oregon, USA). He is an expert in and has published widely on the Munda languages, indigenous Siberian languages, linguistic typology, Native North and South American languages, Eleme (Nigeria) and Burushaski, having worked with speakers of indigenous, endangered, threatened and minority language communities around the globe (including numerous Munda languages). Dr Anderson is also the co-developer of the Language Hotspots theory.

Arun Ghosh is Professor of Linguistics at Burdwan University in West Bengal, India. He is an expert in Bangla and Munda languages. He has worked with speakers of a number of Munda languages, both North Munda and South Munda. In addition to his expertise in Santali, Professor Ghosh is currently working on issues in the grammar of Bonda (Remo) and Didayi (Gtaʔ). He has published extensively, not only on Munda linguistics but also on broader Austroasiatic issues.

Arlo Griffiths is Professor of Sanskrit at Leiden University in the Netherlands. He is an internationally renowned Vedic scholar having produced the first translation of the Paipillad rescension of the Atharvaveda. His interest in Gutob, including several field visits, began a decade ago before his rise to becoming the youngest Chair in Sanskrit at Leiden.

K. David Harrison is Assistant Professor of Linguistics at Swarthmore College, Pennsylvania, USA and Director of Research at the Living Tongues Institute for Endangered Languages. He is the leading recognized international expert on endangered languages. Dr Harrison has worked with speakers of half a dozen Munda languages, both North and South Munda, during field trips in Orissa. Dr Harrison is also the co-developer of the Language Hotspots theory.

Masato Kobayashi is Assistant/Associate Professor of Linguistics at Hakuoh University. Dr Kobayashi has an unparalleled breadth of knowledge of South Asian linguistics. His extreme erudition is attested to not only by his work in Munda linguistics, but also by the fact that he has published extensively in comparative Indo-Aryan linguistics, and in a host of varied topics in Dravidian linguistics.

Ganesh Murmu is Associate Professor and Senior Researcher at the Department of Tribal Languages and Literatures at Ranchi University, Jharkhand, India. A Santal by ethnicity and speech, Dr Murmu brings a unique insight in Kherwarian linguistic analysis to this volume and his many excellent publications continue the great tradition of Munda-speaking linguists, including J. Kullu, P. Kerketta, and Dr R.D. Munda.

Toshiki Osada is Professor at the Research Institute for Humanity and Nature located in Kyoto, Japan. He is the recognized expert on the Kherwarian group of languages,

being an excellent speaker of Mundari. His diverse list of publications in Munda linguistics include lexical materials, field notes on undescribed varieties and theoretically oriented papers on word classes in linguistic analysis, among them works in both Japanese and English. Professor Osada is also a director of the Japanese-led comprehensive project on the Harrappa and Indus civilizations in Pakistan.

Manideepa Patnaik is the recognized expert in both the analysis and description of the Juang language, which she has researched for the past two decades, as well as in the syntax of a number of Munda languages including Sora. Dr Patnaik is currently the Director of the Institute of Indian Languages Literatures and Cultures, Orissa. Her publications and presentations in Munda linguistics have focused primarily on the structure of complex clauses or sentences in languages of South Asia and in Munda languages such as Juang specifically.

John Peterson is Professor of Linguistics at Osnabrück University, having finished his monumental *Habilitationsschrift* on Kharia in 2006. This now sets Kharia apart as the single best described Munda language, in particular in terms of its syntactic features which will be very important to understanding the (pre-)history of this important and somewhat enigmatic Munda language. Professor Peterson's work on Munda has been focused on descriptive and theoretical aspects of Kharia syntax and grammar more widely. He has also published extensively on northern Indo-Aryan languages.

Felix Rau is a PhD student at Leiden University. He is a promising young scholar who already has considerable expertise in Gorum, Desia, Telugu and a range of other languages. His contribution to this volume suggests a sophisticated and nuanced appreciaton of Munda language structure that is advanced for someone of any age. Munda specialists are eagerly awaiting his field-based description of Gorum.

Norman H. Zide is Professor Emeritus at the University of Chicago and is the founder of contemporary Munda linguistics as a scholarly discipline in the West (in India it must be Ramamurti). He has overseen the development of the field from its embryonic state, and initiated the first systematic and only non-Indian (to date) Munda Languages Survey. He has published or written papers in circulation among *cognoscenti* on almost every imaginable topic in Munda linguistics, ranging from Gutob baby talk to comparative Munda phonology, cultural history, echo formation in Gta?, the morphosyntax of verb agreement, and Korku tonal processes, etc.

PREFACE

The Munda Languages volume herein is the result of nearly a decade of work. Although nine years have passed since the chapters were first commissioned from their original intended authors, this was deemed too short a time for some. As originally conceived, the volume was very different. For a variety of reasons the volume has morphed into what it is now, twelve rather different sketches of the major languages and subgroups of the Munda language family, the westernmost representative of the far-flung Austroasiatic linguistic phylum as well as a brief discussion of Nihali/Nahali which may be related to the Munda languages (or may be an isolate language).

It has not been a labour of love for the editor but it has indeed been a labour. Several would-be contributors were unable to offer their work and others were prevented for a wide range of personal reasons from contributing all that they had intended. One chapter changed its author[s] three times, others twice. Indeed, never before has such a collection of works on Munda languages been attempted on this scale (nor is likely to be again for a long time, if ever). Some of the languages in this volume receive here their first comprehensive description, e.g. Gta? or Kera? Mundari.

All in all, this book represents the state of Munda linguistics at the beginning of the twenty-first century. Contributors are taken from all over the world: three from India (including one native Munda-speaking scholar), two from Japan, three from central Europe (Germany/Holland) and three from the United States. To be sure, this includes contributions from practically every researcher currently active in the field of Munda linguistics with only a few exceptions. Each approaches the analysis of the Munda languages from a different background and perspective. Naturally this includes some variation in both notation and terminology, as well as certain points of analysis. For example, palatal nasals may be represented by *ñ* or by *ɲ*, etc. Other conventions, on the other hand, are followed across the chapters including the use of *c* and *j* for the palatal obstruents (stops/affricates) and the IPA symbols for the retroflex (or post-alveolar) series, viz. *ɖ*, *ɽ*, *ʈ*, etc. Note that variation in analysis is seen within a single article, where all parties in co-authored papers may not agree with each other or share the same interpretation of the data. Overall it is believed that the interested reader will derive a wealth of benefits from the materials presented in this volume.

The chapters all follow the same general outline to make comparison easier across them. Some individual chapters deviate from the common pattern of numbered sections slightly, but all sub-headings are clearly marked and the volume is well-indexed so these are mainly minor cosmetic issues.

The editor would like to thank the following people for helping to bring this volume into existence. First, thanks to all contributors for the fine content of their chapters. An especially large thank you must be given to Don Reneau, who as

always has offered me sage advice and help on technical formatting issues, including literally running to the rescue of a contributor who was having insurmountable technical difficulties. Manideepa Patnaik facilitated arrangements for the field trips to India. Mark Eglinton offered his photographic expertise on the 2005 field trip. Sara Foerster helped with a small but gratefully appreciated fraction of the re-keying of contributions by those who did not use Word and/or the SIL IPA fonts. Special thanks are here offered to Ironbound Films who made the 2007 field trip a possibility and the Living Tongues Institute for Endangered Languages for allowing me the opportunity to do this research. I would also like to thank my wife Mary and youngest son Oliver for bearing me during the final months of preparation when these issues were literally driving me insane. I would also like to thank my primary language consultants who shared with me their knowledge and time and patience and without whom the final product would have been impossible and indeed pointless. These include K.C. Naik and C.M. Haibru (Ho), C.M. Singh (Bhumij), S. Dangada-Majhi (Remo), and Opino Gomango and Oruncho Gomango (Sora), among others.

Gregory D.S. Anderson
Salem, Oregon, USA
December 2007

ABBREVIATIONS

-	Affix boundary
.I	First conjugation
.II	Second conjugation
=	Clitic boundary
1	First person
2	Second person
3	Third person
A	Active
A:TEL	Anticipatory telic
ABL	Ablative
ACC	Accusative
ACT	Active
ADD	Additive
ADESS	Adessive
ADJ	Adjective
ADV	Adverbial
AFF	Affective/Affected
AKT	Aktionsart
ALL	Allative
AMB	Ambulative
AN	Animate
ANAPH	Anaphoric Pronominal
ANT	Anterior
ART	Article
ATTR	Attributive
AUGM	Augment
AUTOPOES	Auotpoesis
AUX	Auxiliary
BEN	Benefactive
C:TEL	Culminatory telic
CAP	Capabilitive
CAUS	Causative
CF	Combining form
CLOC	Cislocative
CLSSFR	Classifier
CMPLT	Completive
CNCTV	Connective
CNTR	Contrastive focus
COLL	Collective
COM	Comitative
COMP	Complementizer
COMPL	Completion
CONAT	Conative
COND	Conditional
CONJ	Conjunctive
CONT	Continuative
CONTNG	Contingent action mood
COP	Copula
CR	Correlative
CUST	Customary
CV	Converb
DAT	Dative
DEF	Definite
DEIC	Deictic
DEM	Demonstrative
DEP	Dependent
DESID	Desiderative
DIR	Directional
DISC	Discourse element
DIST	Distal
DL	Dual
DPT	Departive
DS	Different subject
DUBIT	Dubiative
DUR	Durative
EM	*Encyclopedia Mundarica*
EMOT	Emotive 'light verb'
EMPH	Emphatic
EPEN	Epenthetic
EX	Exclusive
EXCES	Excessive
EXCLAM	Exclamatory
EXPR	Expressive
FEM	Feminine
FIN	Finite
FOC	(Restrictive) Focus
FUT	Future
GEN	Genitive

HON	Honorific
HPA	Homo-organic post-alveolar
HUM	Human
IMP	Imperative
IMPF	Imperfect
INAN	Inanimate
INC	Inclusive
IND	Indicative
INDEF	Indefinite
INESS	Inessive
INF	Infinitive
INFER	Inferential
INGR	Ingressive
INS	Instrumental
INTENS	Intensifier
INTERJ	Interjection
INTNSV	Intensive
IPFV	Imperfective
IRR	Irrealis
ITER	Iterative
ITR	Intransitive
LEX	Lexical Verb
LOC	Locative
LV	Loan Verb
M	Middle (Kharia, Santali)
MASC	Masculine
MID	Middle (Gutob)
MID.1	Middle deictic 1
MID.2	Middle deictic 2
MOD	Modal
MODFR	Modifier
N.SFX	*n*-suffix
NARR	Narrative
NEG	Negative
NF	Non-Finite
NHUM	Non-Human
NM	North Munda
NML	Nominalizer
NMLZ	Nominalizer
NONHUM	Non-Human
NPST	Non-past
NP	Noun-Phrase
NUM	Number
OBJ	Object[ive]
OBLQ	Oblique
OPT	Optative
PASS	Passive
PERS	Person
PFV	Perfective
PFX	Prefix
PL	Plural
PLUP	Pluperfect
POL	Politeness
POSS	Possessive
PP	Postposition
PR	Pronominal
PRF	Perfect
PROG	Progressive
PROHIB	Prohibitive
PRON	Pronominal
PROX	Proximal
PRS	Present
PRTCL	Particle
PRTCPL	Participle
PST	Past
PURP	Purposive
Q	Interrogative
QF	Quantifier
QUANT	Quantifying
QUOT	Quotative
RECIP	Reciprocal
REDPL	Reduplication
REP	Repetition (full copy of word)
REL	Relative
RFLXV	Reflexive
RHET	Rhetorical
SBJNCT	Subjunctive
S:ITER	Semel-iterative
SEQ	Sequential converb
SG	Singular
SIMULT	Simultaneous
SM	South Munda
SOC	Sociative
SS	Same Subject
SUB	Subessive
SUBJ	Subject
SUBORD	Subordinate
SUD	Sudden action
SUPERESS	Superessive
SUPERLAT	Superlative
T/A	Tense/Aspect
TAG	Tag-word
TAM	Tense-Aspect-Mood

TEL	Telicity	VBLZ	Verbalizer
TLOC	Translocative	VOC	Vocative
TNS	Tense	W/O	Without
TOTAL	Totality	x1/Y/x2	Y = Infix
TR	Transitive	x1<Y> x2	Y = Infix

This page intentionally left blank

CHAPTER ONE

INTRODUCTION TO THE *MUNDA LANGUAGES**

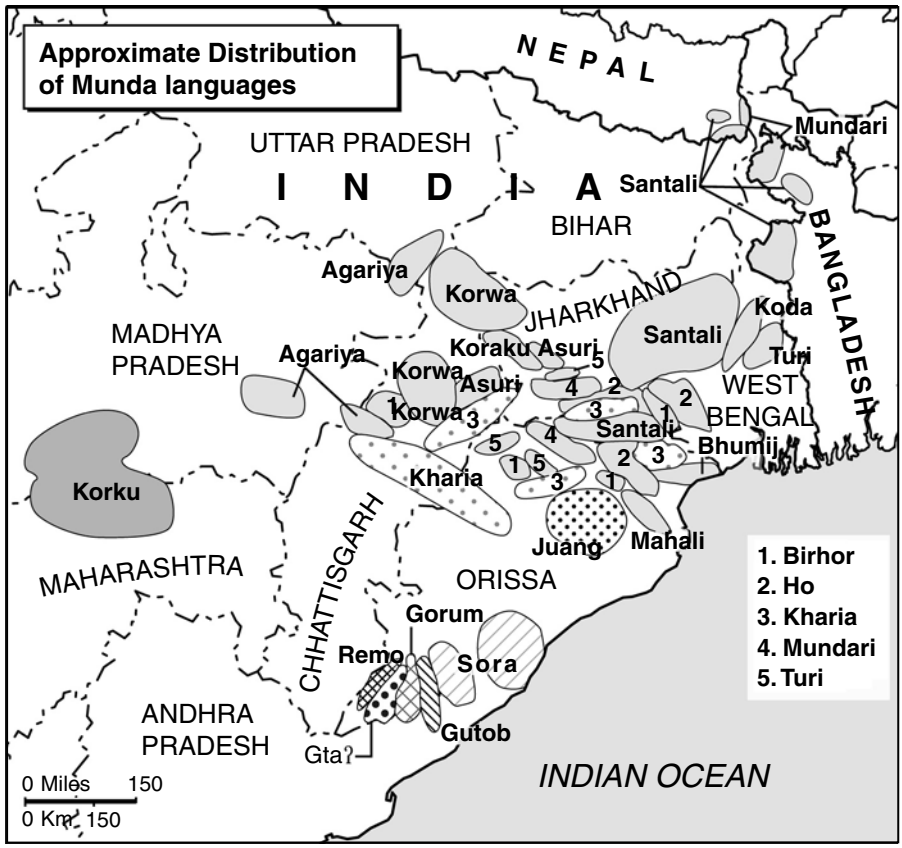
Gregory D.S. Anderson

1 OVERVIEW

The Munda languages are a group of Austroasiatic languages spoken across portions of central and eastern India by perhaps as many as ten million people total. The Munda peoples are generally believed to represent the autochthonous populations over much of their current areas of inhabitation. Originally, Munda-speaking peoples probably extended over a somewhat larger area before being marginalized into the relatively remote hill country and (formerly) forested areas primarily in the states of Orissa and the newly constituted Jharkhand; significant Munda-speaking groups are also to be found in Madhya Pradesh, and throughout remote areas of Chhattisgarh, West Bengal, Uttar Pradesh, Andhra Pradesh, and Maharashtra (see Map 1.1). Of course much of this territory was settled or colonized by the Indo-Aryan speakers and, at an earlier period, by the Dravidian-speakers as well, which concluded about 2,500 years ago or so.

The pre-history of the Munda languages remains obscure. Munda languages constitute the westernmost representatives of the far-flung Austroasiatic linguistic phylum. Two other Austroasiatic groups are found in the present-day territory of India, the Khasi of Meghalaya and the Nicobarese-speaking groups of the Nicobar Islands. The other subgroups of Austroasiatic are all found outside of India, and it is generally believed that the Austroasiatic ancestral language was not to be found in India but rather further to the east. Thus, at some point the ancestors of the Munda-speaking peoples must have migrated westward into the subcontinent. When, how, and by what path they entered India remains a subject of considerable debate. Indeed, it is not even clear that there was a single migration of pre-Munda speakers, but there may have been two or more such movements.

Consensus has not yet been reached on the internal relationships of the Munda languages, but several subgroups have been proposed and some of these appear to be sound. It is hoped that further work in comparative Munda grammar and lexicon may shed light on this issue. The northern-, eastern- and westernmost groups of Munda languages are clearly related and appear to fall into two broad groupings. The first of these is the westernmost Munda language Korku (Zide, this volume-a) which appears to be a sister to the remainder of this subgroup, conventionally labelled Kherwarian. This latter unit is really a large and complex dialect/language chain, the better-known varieties of which are known as Santali (Ghosh, this volume) – by far the largest of the Munda groups – and its close sister languages Mundari (Osada, this volume) and Ho (Anderson *et al.*, this volume). Kherwarian also includes a number of minor varieties as well, for example, Turi, Asuri, Birhor, Bhumij, Korwa, etc. as well as a special quasi-creolized variety of Mundari used by former Kurukh speakers, that is,



MAP 1.1 APPROXIMATE DISTRIBUTION OF MUNDA LANGUAGES

Kera? (Kobayashi and Murmu, this volume). Korku and Kherwarian together are conventionally known as North Munda; this is a secure subgroup within Munda.

The remaining Munda languages are almost only found in the state of Orissa (some Kharia speakers are found in Jharkhand, West Bengal, and Chhattisgarh as well), which appears to be the epicentre of diversity of the family (see Map 1.2). How each of these non-North Munda languages or subgroups (logically known as South Munda in contrast to North Munda) are related to each other remains a topic of considerable debate. Some languages clearly form subgroups, such as Sora (Anderson and Harrison, this volume) and Gorum/Parenga (Anderson and Rau, this volume) in the Sora-Gorum subgroup or Gutob (Griffiths, this volume) and Remo/Bonda (Anderson and Harrison, this volume) in the Gutob-Remo subgroup. The classification of the remaining three languages remains an open question, as does how exactly the non-North Munda languages diversified and developed from the common Proto-Munda ancestral language. These three languages are Kharia (Peterson, this volume), Juang (Patnaik, this volume), and Gta?/Didayi (Anderson, this volume).

The 'traditional' classification of Munda is to be found in Zide (1969) and Zide and Stampe (1968), and may be represented graphically as follows (see Figure 1.1).

MAP 1.2 APPROXIMATE DISTRIBUTION OF MUNDA LANGUAGES OF ORISSA

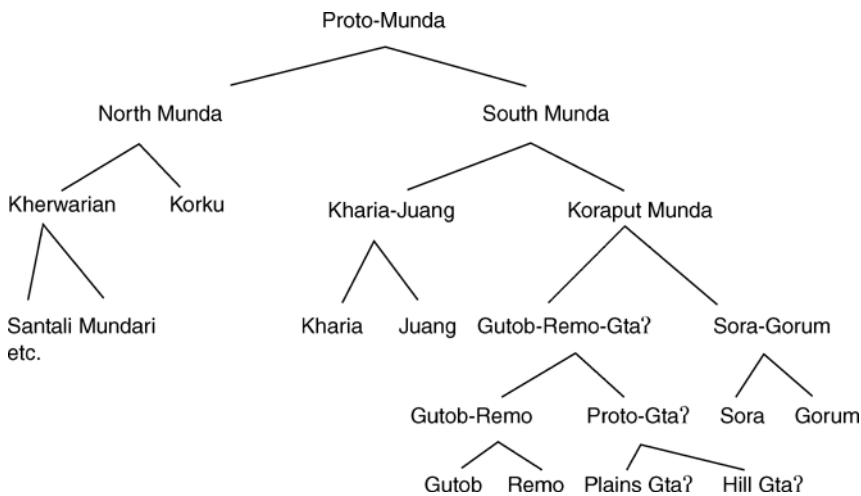


FIGURE 1.1 TREE DIAGRAM DETAILING 'TRADITIONAL' CLASSIFICATION OF MUNDA LANGUAGES

Anderson (1999) is a first recent attempt to rethink this traditional classification. He modified the tree as seen in Figure 1.1 only slightly by eliminating the Koraput Munda node and having South Munda branch directly into three daughter groups, still keeping Kharia-Juang and Gutob-Remo-Gta? intact (see Figure 1.2).

A more radical revision was proposed by Anderson (2001), which adopted a slightly different approach, whereby only the clearly obvious subgroups within South Munda were recognized as such (see Figure 1.3). He proposed an early dialect continuum in which various languages shared certain features (some retentions and some innovations) but did not form a taxonomic unit *per se*. Such a chain has been called by Malcolm Ross (e.g. 1996) 'a linkage' in the literature on Austronesian and Papuan languages. This revision is largely motivated by the curious, non-archaic, and multifaceted parallels shared between Kharia and Proto-Gutob-Remo which the traditional understanding of South Munda could not explain.

Whether and/or how Munda is related to the enigmatic Nihali/Nahali language of Madhya Pradesh is a complicated question, but for an introduction to the issues, see Zide (this volume-b).

It is surprising that nothing in the way of quotations from a Munda language turned up in (the hundreds and hundreds of) Sanskrit or middle-Indic texts. There

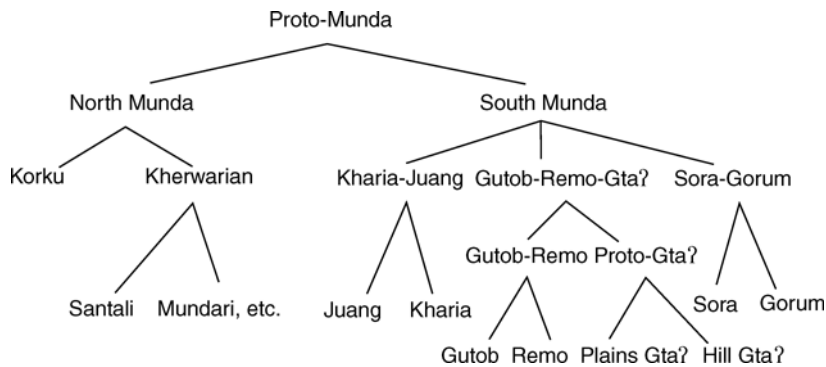


FIGURE 1.2 TREE DIAGRAM DETAILING REVISED TRADITIONAL CLASSIFICATION OF MUNDA LANGUAGES (ANDERSON 1999a)

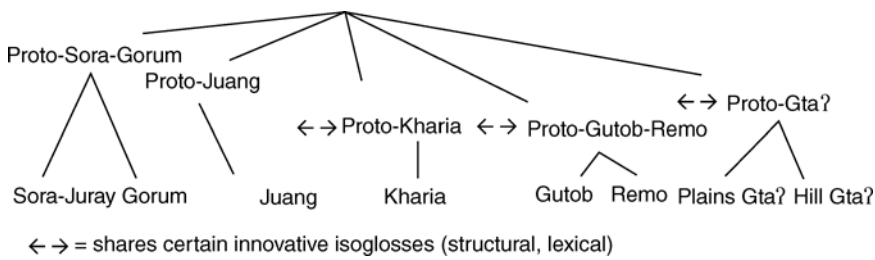


FIGURE 1.3 TREE DIAGRAM OF CLASSIFICATION OF PROTO-SOUTH MUNDA LANGUAGES AS REVISED BY ANDERSON (2001)

is also a surprising lack of borrowing of names of plant/animal/bird, etc. into Sanskrit (Zide and Zide 1976). Much of what has been proposed for Munda words in older Indic (e.g. Kuiper 1948) has been rejected by careful analysis. Some possible Munda tribal names have been proposed, for example, *Savara* (Sora) or *Khara*, but ethnonymy is notoriously messy for the identification of language groups, and a single ethnonym may be adopted and used for linguistically rather different or entirely unrelated groups.

Despite being spoken in a country with a written tradition over two millennia old, the Munda languages remained in total obscurity until the middle of the nineteenth century. Throughout the twentieth century, with a few notable exceptions, the minor Munda languages were quietly disappearing while the intellectual community (both in India and abroad) showed no concern. Because the so-called Kol(arian) languages, as this group was frequently known in the past, were spoken by 'backward tribals', they received little attention from early scholars. Various Christian missionaries have provided us with many of the early works on the Munda languages.

The first known reference to data from a Munda language may be Voysey (1821, published 1844). Before Cust (1878, 1884), most of the mention of the Munda languages comes from vocabularies and early primers, for example, Tickell (1840a/b), Hodgson (1848), Phillips (1845), Phillips (1852, Santali grammar), Campbell (1866, Santali vocabulary), Puxley (1868, Santali vocabulary), Das (1871, Mundari grammar), Skrefsrud (1873, Santali grammar), Coates (1875, Chota Nagpur vocabulary) or Cole (1879, Santali article), Pendercast (1881, Sora vocabulary). During the 1890s Rev. de Smet's *Rudiments of Mundari Grammar* (1891), Crooke's (1892) Korwa vocabulary, Banerjee's *Kharia Grammar* (1894), Hoffmann's early work (1893), and part 1 of Campbell's *Santal Dictionary* (1899) were published.

The first decade of the twentieth century witnessed a relative boom in Munda linguistics; for example, the second and third parts of Campbell's dictionary appeared (1900, 1903). Other important publications of this era were Nottrott's *Kol Grammar* (1905), Hoffman's Mundari works (1903, 1909) and Cole's Santali study (1906). This period had also witnessed Grierson's *Linguistic Survey of India* (volume 4) and various other works and Schmidt's groundbreaking study establishing the Austroasiatic connection of the Munda languages (1906). Burrows' *Ho Grammar* (1915), various studies by Roy, Przyluski and especially P. Bodding's Santal materials were highlights of the 1910s and 1920s, as well as the ethnographically interesting 'Tea District' handbooks, with their bizarre phrase-book qualities that were produced for Sora, Kharia, and Mundari, in the wake of an increasing Munda-speaking migrant labour force mobilized to work in the tea plantations of Assam and Darjeeling.[1] The period from 1930s through 1940s featured the works of the French Austroasiaticists Haudricourt and de Hevesy, and the Dutch Indologist F.B.J. Kuiper; however, two studies of this era stand far above the others in their impact on Munda linguistics – Hoffmann's seminal *Encyclopedia Mundarica* (1930–1950) and Ramamurti's landmark *Manual of the Sora (Savara) Language* (1931). The 1950s produced works by MacPhail and Biligiri, H. Maspero in the French Austroasiaticist tradition and Pinnow's Kharia and Juang materials.

The 1960s and 1970s witnessed another boom in Munda linguistics. Pinnow (1966) is the first comprehensive study on the verb in the Munda languages. Written in 1960, it lacks knowledge of the existence of Gta?. Norman Zide, the pre-eminent Western scholar in the field of the emergent discipline of Munda linguistics, led a group of linguists (including A. Zide, D. Stampe, K.C. Bahl, and R.D. Munda at

the University of Chicago and other linguists from a range of institutions, including F. Fernandez, S. Starosta, D. Matson, R. King, etc.) under the auspices of the Munda Language Project, which produced a number of dissertations and smaller works in this period. Indian linguists involved in Munda studies in this period include B. Das, K. and B.P. Mahapatra and S. Bhattacharya. In the next two decades, in addition to a variety of publications by N. Zide, A. Zide and S. Starosta, P. Donegan and D. Stampe, Munda languages have occasionally garnered interest from linguists pursuing a range of typologically or theoretically oriented lines of research, including Masica (1976; various South Asian areal features), and Mithun (1984; with brief discussions of incorporation in some South Munda languages).

In the 1990s and early twenty-first century, a new generation of linguists has undertaken an intensive study of the Munda languages. In addition to Norman Zide, still active in this period, the names of Toshiki Osada, John Peterson, Arun Ghosh, Greg Anderson, Manideepa Patnaik, Ganesh Murmu, K.S. Nagaraja, Felix Rau, Masato Kobayashi, and N. Ramaswami have appeared in connection with a range of descriptive, typological, and comparative historical studies of the Munda language family. Linguists approaching Munda data from a theoretical or South Asian areal perspective from the last two decades include Abbi (1992, reduplication), Odden (1987, theoretical aspects of Gta? phonology), Hook (1991, on complex verb structure), and Sadock (1991, syntactic/theoretical aspects of Gta? incorporation).

A number of Munda-speaking scholars have actively participated in the linguistic analysis and study of the Munda languages such that an indigenous scholarly tradition may be spoken of. These include the Santals Ganesh Murmu and R. Murmu, D. Sahu, Kharia J. Kullu, and P. Kerketta to name but a few.

Most Munda languages remain unwritten or have only fledgling literary varieties. Given the pressure to have a script of one's own to be considered a 'real' language prevalent in the South Asian area (Zide 1996, 2000), it comes as no surprise that there are three or four of the Munda languages which have had indigenous script traditions in the twentieth century. Of these, only the Warang Chiti script of the Ho appears to have any chance of gaining acceptance (see Anderson *et al.* [this volume] on Warang Chiti and Ghosh [this volume] for a discussion of the seeming failure of the Ol Ciki script of the Santals, also Pinnow 1972).

What will researchers find interesting about the Munda languages and what remain the outstanding issues in the descriptive analysis of these languages? Little data on the Munda languages are in wide circulation among linguists who may be inclined to know about them. This volume will serve as a first step in the direction of bringing to light some of the noteworthy features of the individual Munda languages and the family as a whole. All chapters in this volume more or less follow the same general outline, and are largely weighted to morphological structures in the languages. Nevertheless, researchers on such diverse topics as reduplication, noun incorporation, articulatory phonetics, numeral systems, agreement morphosyntax, complex predicate structure, or nasalization will all find some aspect of some Munda language intriguing, stimulating or even challenging analysis.

Among the most interesting of the linguistic phenomena to be found in Munda languages may be included the highly elaborated systems of demonstratives found in many Munda languages (see, for example, the discussion of these systems in the chapters on Santali, Mundari, or Gorum in this volume). Munda vowel and consonant systems can be quite complex, with different register and secondary articulatory features, many of which are still now in need of description. Another topic of

considerable interest are the elaborate and intersecting systems of voice/valence/transitivity, person-marking, and tense/aspect that characterize Kherwarian verbal systems. Further, the highly elaborated system of noun incorporation found in Sora push the limits of our understanding of such constructions from a theoretical perspective. The highly developed systems of reduplication and expressive formation that characterize most Munda languages also bear mention here. Finally, the interaction of tense/aspect marking and negative operators in negative formations in South Munda Gutob stand out among the most complex of such systems known.

As for what topics remain for future analysis, it can be said without reservation, that almost all aspects of every Munda language require more analysis before we have an adequate consensus understanding of even their basic features. That said, it is clear from a comparison of the various chapters that syntactic issues and phonetic analysis are in desperate need of further systematic investigation. Studies on topics in the semantics and discourse of Munda languages are practically non-existent. Comprehensive comparative study has not been really possible up to this point either on the lexicon or the grammar (though many excellent preliminary attempts have been made, particularly by Pinnow, N. Zide, A. Zide, Stampe, and Anderson), so a more thorough and comprehensive investigation into most historical linguistic issues in Munda also remains a goal for the future. This will aid considerably in the comparison of Munda with its Austroasiatic sister languages, as well as begin to gain a better understanding of the (linguistic) pre-history of South and Southeast Asia.

I will not give a brief introduction to the content of each chapter pointing out their strengths and weaknesses here as is typical in collected volumes such as this. I leave it to the interested readers to draw their own conclusions. Suffice it to say, that while this volume represents a monumental leap forward in the advancement of Munda linguistics, there is still much ground to cover before an adequate description of the languages will have been achieved. However, as the reader will see, these languages have much to contribute to our understanding of such diverse topics ranging from noun incorporation, transitivity, or the very nature of human language itself to the pre-history and contemporary sociolinguistics of India.

NOTES

* Thanks to Norman Zide for some suggestions for this chapter. All errors remain my responsibility.

1 The sample sentences in the Tea District publications include such useful phrases as Sora *barasui badimaranji raptiurungtaaji pa?* 'do you have the chance to get other coolies' *pantagooleengan asoongdoong!* 'you must not defecate on the road!' and *unta badimaran uggadong* 'do not abuse that coolie!' (38–39, 89).

REFERENCES

Abbi, Anvita (1992) *Reduplication in South Asian Languages*, Delhi: Allied Publishers.

Anderson, Gregory D.S. (1999) 'A new classification of the Munda languages: Evidence from comparative verb morphology', Paper presented at 209th meeting of the American Oriental Society, Baltimore, MD.

——(2001) 'A new classification of Munda: Evidence from comparative verb morphology', *Indian Linguistics*, 62:27–42.

Anderson, Gregory D.S. (2001) (this volume, Chapter 13) 'Gta?'.
Anderson, Gregory D.S. and K.D. Harrison (this volume-a, Chapter 11) 'Remo (Bonda)'.
——(this volume-b, Chapter 7) 'Sora'.
Anderson, Gregory D.S., T. Osada, and K.D. Harrison (this volume, Chapter 5) 'Ho and the other Kherwarian languages'.
Anderson, Gregory D.S. and F. Rau (this volume, Chapter 8) 'Gorum'.
Banerjee, G.C. (1894) *An Introduction to the Kharia Language*, Calcutta: Bengal Secretariat Press.
Burrows, Lionel (1915/1980) *Ho Grammar with Vocabulary: An Eastern Himalayan Dialect*, New Delhi: Cosmo Publication (originally in Calcutta: Catholic Orphan Press).
Campbell, Alexander (1899, 1900, 1903) *A Santali–English Dictionary*, 3 parts, Pokhuria: Santal Mission Press.
——(1905) *English–Santali; Index to Santali–English Dictionary*, Pokhuria: Santal Mission Press.
Campbell, G. (1866) 'The ethnography of India (Appendix 2: Santal vocabulary)', *Journal of the Asiatic Society of Bengal*, 35(2):1–52.
Coates, J.M. (1875) *Vocabulary of Seven Languages and Dialects of Chotanagpore*, Calcutta: Bengal Secretariat Press.
Cole, Frederick T. (1879) 'Lists of words and phrases with their Santali equivalents', *Indian Antiquary,* 8:194–196.
——(1906) *Santali Primer*. Benegaria: Santal Mission Press.
Crooke, William (1892) 'Vocabulary of the Korwa language', *Journal of the Royal Asiatic Society*, 61(1):125–128.
Cust, R.N. (1878) *A Sketch of the Languages of the East Indies*, London: Trubner.
——(1884) 'Grammatical note and vocabulary of the language of the Korku: A Kolarian tribe in Central India', *Journal of the Royal Asiatic Society of Great Britain and Ireland* (*new series*), 16:164–179.
Das, Rakal (1871) 'An introduction to the Mundari language', *Journal of the Asiatic Society of Bengal*, 40:46–66.
De Smet, Rev. (1891) *Rudiments of a Mundari Grammar*, Calcutta: Catholic Orphan Press.
Donegan, Patricia J. (1993) 'Rhythm and vocalic drift in Munda and Mon-Khmer', *Linguistics in the Tibeto-Burman Area*, 16(1):1–43.
Donegan, Patricia J. and David Stampe (1983) 'Rhythm and the holistic organization of language structure', in *The Interplay of Phonology, Morphology, and Syntax*, Chicago: Chicago Linguistic Society: 337–353.
Fernandez, Frank (1968) *A Grammatical Sketch of Remo: A Munda Language*, PhD dissertation, University of North Carolina.
Ghosh, A. (this volume, Chapter 2) 'Santali'.
Grierson, George A. (1906) *Linguistic Survey of India. Vol IV: Munda and Dravidian*, Calcutta: Office of the Superintendent of Government Printing.
Griffiths, A. (this volume, Chapter 12) 'Gutob'.
Hodgson, Brian H. (1848) 'The aborigines of Central India', *Journal of the Asiatic Society of Bengal*, 17(2):550–558.
Hoffmann, Johan (1893) *A Mundari First Reader*, Calcutta: Bengal Secretariat Press.
——(1903) *Mundari Grammar*, Calcutta: Bengal Secretariat Press.
——(1909) *A Mundari Grammar with Exercises*, Calcutta: Catholic Orphan Press.
——(1930–1950) *Encyclopedia Mundarica*, 15 volumes, Patna: Government Superintendent Printing.
Hook, Peter. E. (1991) 'The compound verb in Munda: An areal overview', *Language Sciences*, 13(2):181–195.

Kobayashi, M. and G. Murmu (this volume) 'Kera? Mundari'.
Kuiper, F.B.J. (1948) *Proto-Munda Words in Sanskrit*, Amsterdam: Noord-Hollandsche Maatchappij.
Masica, Colin (1976) *Defining a Linguistic Area*, Chicago: University of Chicago Press.
Mithun, Marianne (1984) 'The evolution of noun incorporation', *Language*, 4(60): 847–894.
Nottrott, Alfred (1905) *Kol Grammar*, (translation by A.G.F. Fitzgerald and Paul Wagner), Ranchi: G.E.L. Mission Press (original edition 1882, *Grammatik der Kolh-Sprache*, Gütersloh: C. Bertelsmann).
Odden, David (1987) 'Arguments against the vowel plane in Gta?', *Linguistic Inquiry*, 18(3):523–529.
Osada, T. (this volume, Chapter 3) 'Mundari'.
Patnaik, M. (this volume, Chapter 9) 'Juang'.
Pendercast, M.H. (1881) 'Short vocabulary of the Savara language', *Journal of the Royal Asiatic Society*, 13:425–428.
Peterson, J. (this volume, Chapter 10) 'Kharia'.
Phillips, Jeremiah (1845) *A Santali Primer*, Calcutta: School Book Society.
——(1852) *An Introduction to the Santali Language*, Calcutta: School Book Society.
Pinnow, Heinz-Jürgen (1959) *Versuch einer historischen Lautlehre der Kharia-Sprache*, Wiesbaden: Otto Harrassowitz.
——(1960) *Beiträge zur Kenntnis der Juang-Sprache*, Unpublished manuscript, Berlin.
——(1966) 'A comparative study of the verb in the Munda languages', in N. Zide (ed.) *Studies in Comparative Austroasiatic Linguistics* (Indo-Iranian Monographs vol. 5), The Hague: Mouton, 96–193.
—— (1972) 'Schrift und Sprache Lako Bodras', *Anthropos*, 67:822–857.
Pinnow, Heinz-Jürgen, J. Kullu and P. Kerketta (1965) 'Der Wertiger und andere Texte', *Indo-Iranian Journal*, 9(1):32–68.
Puxley, Edward L. (1868) *A Vocabulary of the Santali Language*, London: W.M. Watts.
Ramamurti, G.V. (1931) *A Manual of the So:ra: (or Savara) Language*, Madras: Government Press.
Ramaswami, N. (1992) *Bhumij Grammar*, Mysore: Central Institute of Indian Languages.
Ross, Malcolm (1996) 'On the genetic affiliations of the Oceanic Languages of Irian Jaya', *Oceanic Linguistics*, 35(2):258–271.
Sadock, Jerrold (1991) *Autolexical Syntax*, Chicago: University of Chicago Press.
Schmidt, Pater P.W. (1906) 'Die Mon-Khmer Völker: Ein Bindeglied zwischen Völkern Zentralsiens und Austronesiens', *Archiv der Anthropologie, Braunschweiger*, 33[n.s. 5]:59–109.
Skrefsrud, Lars Olsen (1873) *A Grammar of the Santhali Language*, Benares: Medical Hall Press.
Starosta, Stanley (1967) Sora Syntax: A Generative Approach to a Munda language, PhD dissertation, University of Wisconsin.
Tickell, S.R. (1840a) 'Grammatical Construction of the Ho language', *Journal of the Asiatic Society of Bengal*, 9:997–1007.
——(1840b) 'Vocabulary of the Ho Language', *Journal of the Asiatic Society of Bengal*, 9:1063–1090.
Voysey, H.W. (1844) 'Vocabulary of Goand and Cole words', *Journal of the Asiatic Society of Bengal*, 13(1): 19–23.
Zide, N.H. (1969) 'Munda and non-Munda Austroasiatic languages', in Thomas Sebeok (ed.) *Current Trends in Linguistics*, The Hague: Mouton, 5:411–430.

Zide, N.H. (1996) 'Scripts for Munda languages', in Peter Daniel and William Bright (eds) *The World's Writing Systems*, New York: Oxford University Press, 612–618.

——(2000) 'Three Munda Scripts', *Linguistics of the Tibeto-Burman Area*, 22:199–232.

——(this volume-a, Chapter 6) 'Korku'.

——(this volume-b, Chapter 14) 'On Nihali'.

Zide N.H. and Stampe, D.L. (1968) 'The position of Kharia-Juang in the Munda family' in B. Krishnamurti (ed.) *Studies in Indian Linguistics* (Professor M.B. Emeneau's śaśtipurti volume), Poona: Centre of Advanced Study in Linguistics, Deccan College, 370–377.

Zide, N.H. and A.R.K. Zide (1976) 'Proto-Munda cultural vocabulary: Evidence or early agriculture' in P.N. Jenner, C. Thompson, and Stanley Starosta (eds) *Austroasiatic Studies, Part II*, Honolulu: University of Hawaii Press, 1295–1334.

CHAPTER TWO

SANTALI

Arun Ghosh

1 THE LANGUAGE AND ITS SPEAKERS

1.1 Linguistic type

Santali is a Munda language with a suffixing, agglutinating, and a basic SOV structure. The most notable characteristic of the language is the weak distinction between noun and verb, the addition of verbal suffixes that will turn any lexeme into a verb, and case-markers, enclitic definitives, and number markers into a noun. In the present study, therefore, the word classes have been postulated on the basis of morpho-syntactic criteria rather than lexical criteria alone.

The consonant inventory consists of five categories of plosives distinguished with regard to five places of articulation (labial, alveolar, post-alveolar, palatal, and velar), a nasal corresponding to each, one trill, one flap, one lateral, two fricatives, and two semivowels. There are eight vowels, without any significant length distinction. Words are predominantly dissyllabic with stress on the second syllable in dissyllabic and multisyllabic words, and on the root vowel if it is monosyllabic. Diphthongs do occur with rising and falling varieties. Consonant clusters are rare. As to the syllable structure the language has a predilection for CV structure closely followed by CVC. One notable feature of the language is that nasals can form the nucleus, found nowhere else among Indian languages apart from Munda.

Grammatically speaking there are classes of nominals (noun and pronoun), verb, demonstrative, adjective, adverb, and particle. The language has a gender agreement system. Nouns and demonstratives show a two-way gender distinction – animate and inanimate. Furthermore, some nouns show an overt masculine–feminine distinction in their forms. Nouns are declined in three numbers – singular, dual, and plural. Pronouns show forms for all the three numbers in all three persons with an inclusive–exclusive distinction in the first person. The third personal pronoun is actually derived from the distal demonstrative root. Use of the anaphoric third personal form is not very common. The demonstrative is divided into three classes – simple, interrogative, and indefinite. Each class shows an animate–inanimate distinction. Case is affected by the use of suffixes and postpositions. Whereas sociative and ablative are marked by postpositions, the genitive, comitative, instrumental-locative, allative, and locative are marked by suffixes. Nominative is marked on the verb as transitive subject, intransitive subject, and transitive object. Case is divided into core and peripheral. Whereas the core agrees with the verb in the form of pronominal argument (subject, direct object, and indirect object), the peripheral is marked on the nominals and does not have any agreement with the verb. The root morphemes, derived forms, and phrases which serve as attributes in endocentric attributive constructions and as predicate complements in subject–predicate complement sentences are grouped as adjectives. Adjectives do not agree in number or in gender, except

in a few borrowed items, and in the case of qualified nouns. Adjectives are not inflected to show degrees of comparison. To obtain degrees of comparison postpositions are added to the word with which something is to be compared. Numerals are generally found as quantifiers combined with classifiers. The quantifiers are used singly when they are used in enumerating human beings in the indefinite, whereas in counting human beings in the definite and non-human beings and inanimate objects, the quantifiers are used combined with classifiers. Adverbs are modifiers irrespective of the position they occupy in the sentence.

The verb is defined by the fact that it takes TAM suffixes with or without the markers for the active and neutral, personal terminations, the marker for finiteness of action, and the gerundial suffix. Words which denote actions, events, or conditions are not the only ones treated as verbs, but any word is here treated as such provided that it takes verbal suffixes. For finiteness of action it requires the finite marker /*a*/.

The TAM suffixes have two forms depending on whether it is active or middle – one used on the verb in transitive constructions with the other being employed on verbs in intransitive constructions. The ambitransitive roots take on active as well as middle suffixes. Verbal roots fall into three classes: intransitive only, transitive only, and ambitransitive.

Verbal stems are divided into causative, reciprocal, benefactive, passive/reflexive, mediopassive, iterative/intensive, and compound, and morphological processes are employed to form stems including prefixation, infixation, suffixation, reduplication, and compounding. Except the particles /*kan*/ and /*taẽkan*/ for the present and past tenses respectively there are no separate tense markers. There is a concept of time dimension in the present, past, and future tenses and the aspect suffixes along with their own functions denote time as well.

On the syntactic level, generally the head follows its determiner. In the simple sentence then word order is SOV. In complex sentences, conjunctions, coordinate clauses, and the subordinate clauses perform nominal or attributive function. Sentence modification is achieved through a set of particles. Polar questions can be shown by a marked intonation pattern.

1.2 Name of the language and the tribe

The name Santali, the language, is derived from the ethnic name *Sãotal*, the Anglicized version being Santal. From *Sãotal* the neighbouring non-Santals use the name *Sãotali* for their language. The Santals themselves call their language *hɔɽ* or *hɔɽ rɔɽ*. Being more analytical they attach *rɔɽ* 'language, speech' to *hɔɽ* 'Santal man', hence 'language of the Santals'. The language is also sometimes returned under the name *mãjhi bhasa* 'language of the Majhis'. In North Bengal, in the districts of Murshidabad, Malda, Dinajpur, etc. the language is known as *jaŋli* or *pahaɽia*. In South Bengal, particularly in the districts of 24-Parganas (north and south) and Howrah, and in Orissa the language is known as *ʈhar*. In Bihar it is known as *parsi* 'foreign'. The name *Sãotal,* is derived, in turn, from *Sāmanta-pāla*[1] 'dweller of the frontier' and is used to name the tribe by the non-Santals, particularly the Bengalees. L.O. Skrefsrud[2] tried to derive the name from *Sãot*, a place in Midnapur in West Bengal where the Santals were supposed to have been settled in remote antiquity. There is still a place called *Sãot* in Midnapur district where Santal habitation is common. The Santals call themselves *hɔɽ* 'man' and the name *Santal* is only used by those who are Christianized. From the term *maɲjhi* or *mãjhi* 'village headman' they

also call themselves *maɲjhi* or *mājhi* when asked about their caste. Being the oldest ethnic stock in India they are also known as *ādivāsi* 'those who have been living in the land from the beginning.' S.K. Chatterji[3] attaches great importance to this term for them remembering the contribution of the Santals in the evolution of Indian life and culture.

1.3 Genetic affiliation of Santali

Santali belongs to the Kherwari group of the North Munda sub-family of the Munda family which is, in turn, a section of the eastern group of the great Austroasiatic family of languages. Pinnow (1959)[4] offered a classification of the Austroasiatic languages into two main branches – West Obergruppe, consisting of the Munda languages and Naha-li[5] and Ost Obergruppe including the rest of the Austroasiatic languages. Norman H. Zide divides the Munda languages into two main groups – South Munda (SM) and North Munda (NM). The eastern (i.e. Kherwari) and the western (i.e. Korku) form the North Munda branch. Santali is separated off as a special branch of Kherwari. South Munda, on the other hand, consists of Koraput Munda (KM) and Central Munda (CM). Koraput Munda consists of Sora-Gorum (SG) group and Gutob-Remo-Gtaʔ (GRG). Central Munda consists of Kharia and Juang.[6] Bhattacharya[7] conforming to the same geographical classification puts forward another classification based on morphological criteria. He proposes a two-way division – Lower Munda, consisting of the three extreme southern Munda languages, ḍiḍeyi (Gtaʔ), Bonda (Remo), and Gutob spoken in Koraput and Malkangiri districts of Orissa, bordering Andhra Pradesh with Upper Munda covering the rest. As the three southernmost languages included in Lower Munda do not show pronominal incorporation in the verb, differing from the rest of the Munda languages in respect to genitive marker and by not having dual number, they are branched off from the rest of the Munda languages.[8] But whatever the situation is, the position of Santali remains the same. It has been branched off as a special member of the northeastern group possibly because it has reached a higher stage of development than any other language of the same group and also preserves the peculiar linguistic features of Munda more faithfully than the rest.

1.4 Number of speakers and locale of concentration

The total population of the Santals is 6,050,000[9] in India, Bangladesh, Bhutan, and Nepal, of which 5,959,000 are in India, 157,000 in Bangladesh, and 33,332 in Nepal. The exact figure for Bhutan is not available. The main concentration of the Santals is in India with scattered settlements in Bangladesh, Nepal, and Bhutan. They are also found all over the world wherever they have found job opportunities. The classified data of 2001 census is for all the states of India, except for West Bengal, which is not available yet. In West Bengal it is estimated to be 2,280,540.

The distribution of the Santals in different states of India, according to the 1981 census, is shown in Table 2.1.

As it appears from the census data the Santal population is spread over a large strip of land covering almost the whole of India. The most compact area of concentration is the western part of West Bengal, the southern portion of erstwhile Bihar, now Jharkhand, the areas of Bihar adjacent to Jharkhand, the northeastern districts of

TABLE 2.1: DISTRIBUTION OF SANTALI SPEAKERS IN INDIA

State	Total	Male	Female
Andhra Pradesh	50	30	20
Andaman & Nicobar Islands	7	3	4
Arunachal Pradesh	484	282	202
Bihar (including Jharkhand)	2,161,032	1,087,820	1,073,212
Chandigarh	10	8	2
Delhi	126	69	57
Goa, Daman & Diu	1	1	—
Gujarat	40	30	10
Haryana	10	5	5
Himachal Pradesh	7	7	—
Jammu & Kashmir	28	26	2
Karnataka	45	30	15
Madhya Pradesh	715	445	270
Maharashtra	110	64	46
Manipur	351	348	3
Meghalaya	212	199	13
Mizoram	2,301	2,278	23
Nagaland	1,100	1,080	20
Orissa	529,574	264,854	264,720
Punjab	10	10	—
Rajasthan	52	37	15
Sikkim	16	14	2
Tripura	3,518	1,812	1,706
Uttar Pradesh	282	257	25
West Bengal	1,632,440	822,973	809,467

Orissa and Assam. The data of 1981 did not give figures for Assam, as census could not be conducted due to political turmoil. According to the 1971 census the figure goes to 86,303.[10] So far as the Jharkhand is concerned the figure is not available, as the state was not formed during the last census. The figure for Bihar in Table 2.1 can be read as that of Jharkhand, as the state of Jharkhand was constituted of the southern portion of Bihar which houses almost the whole of the tribes. If the data are represented in a map the areas just described will form a compact area, comprising western parts of the districts of Birbhum and Burdwan, the Sadar subdivision[11] of Bankura, Jhargram subdivision of West Midnapur, Purulia in West Bengal, the whole of Jharkhand, especially the Santal Parganas, Hazaribagh, Singbhum and Dhalbhum districts, south of Bhagalpur and Munghyr in Bihar, and Balasore, Mayurbhanj and Keonjhar in Orissa. In Assam the Santal settlements are confined to the tea gardens only, where they were brought as labourers. The process of shifting is now transformed into the process of migration for job opportunity in the tea gardens. They are also sparsely distributed in the northern districts of West Bengal,[12] in the districts of South 24-Parganas, Hooghly, Howrah and so on in south Bengal where they are employed as day-labourers. The Santals settled at Rajsahi, Rangpur, and Chattagram in Bangladesh are said to be immigrants, migrated from the other part of the river Padma, that is, India at different periods of history. But considering their concentration in Nepal and Bhutan, and going by the different theories of their origin and migration it may be conjectured that they are sons of the soil and that

others found further west in India migrated for better living.[13] The hypothesis may be strengthened by their presence in Nepal and Bhutan.

The Santals are the most numerous among the tribes who speak Munda. In the western fringe of West Bengal, north Orissa and Jharkhand, normally, they muster very strong. This helps them maintain group solidarity and preserve their language and culture much better than elsewhere. The Santals now remaining in other places are nothing but scattered masses floating here and there, and in the process they are all but melted with other dominant cultures in the region. The greater part of their substance has already commingled in the fluid around them, the remainder is saturated with it, and it is only in the very kernel and inner centre of the largest lumps that something like the pure original substance is to be sought.

1.5 Bilingualism, processes of Aryanization and nativization

The proportion of the population retaining the language decreases as the tribe spreads over to industrial areas, nearer to towns and cities. It is also affected by the spread of education among the younger generation. The more they are educated, the more they are oblivious to their own language and drawn nearer to neighbouring languages of the Indo-Aryan group, say, Bengali in West Bengal, Hindi in Jharkhand and Bihar, Oriya in Orissa, and Assamese in Assam. There are, of course, two reasons behind this attitude; first, living by the side of the majority communities and in mixed localities they have to be bilinguals, as there is no other alternative. Over a period of time they are more with the majority language than their own. Being constantly in an alien system they automatically adapt. Second, retention of a language also has a socio-economic background. In the job market the language has less potential than the neighbouring majority languages. So far as the prestige factor is concerned, for them it has none, since until recently the language was not recognized in the eighth schedule of the Constitution of India. So there were and still are ample reasons for leaving their own language and welcoming the neighbouring one. In West Bengal, as a whole, 55% of the Santals speak their own language but in the district of 24-Parganas the proportion falls to 10%. The total number of bilinguals among the Santals is 1,501,638, being 34.66% of the total.[14] Whereas the national average of bilingualism is 13.34%, it is more than 30% in the case of tribal population. Obviously, necessity for knowing the common language of the area is very great in the case of tribes who speak a language which is not a majority one. As already said, being a minority-language community, they cannot help speaking the dominant language of the region where they stay. In every-day interaction, they have to come across a group of people who belong to a different speech community, namely the Bengali-speech community in West Bengal, the Mundari and the Hindi-speech community in Jharkhand, the Hindi-speech community in Bihar, the Oriya-speech community in Orissa, and the Assamese-speech community in Assam. It is very common then that the local languages will influence their own language in some way or another. This way Santali is, to some extent, modified in phonology, morphology, and lexicon. Syntax is not much altered.The lexical elements of the dominant languages are taken into Santali either through a process of nativization or just as they are. The connotation of some lexical elements is changed under the influence of the neighbouring languages.

New postpositions and particles are borrowed and nativized following the phonotactic rules of the language. The extent of influence on Santali grammar and lexicon may be clarified with certain examples.

Being in contact with the neighbouring world they start adopting its language system along with many words being Santali-equivalent; for example, *ləḍu* 'sweet-meat', *miṯhəi* 'sweet', *dal* 'beans', *caole* 'rice', *pənəhi* 'shoes' have been incorporated into their system. As professions such as shoemaker, blacksmith, and carter do not belong to the Santal society, the corresponding terms like *muci* 'shoemaker', *kamar* 'blacksmith', and *gaḍvan* 'carter' are borrowed from the neighbouring language system. Similarly, for trades not belonging to their system, they have borrowed trade-related terms from the neighbouring language system; for example, *mal* 'goods', *cij* 'sample', *asbab* 'furniture', *bajar* 'market', *khərca* 'expenditure', and so on. In time calculation the Aryan influence is noticeable; thus *ghəɽi* 'a while', *din* 'day', *bɔchɔr* 'year', *cirəkal* 'a long time'. Even though they have their own village administration terms like *maɲjhi* 'village headman' (also *mãjhi* 'id.') (cf. Skt. *madhya*), and *pancayat* 'village council' are borrowed by them from the neighbouring Indo-Aryan languages to accommodate themselves in the government-controlled Panchayat system. Even in the household Aryan influence is noticeable. The inner apartment is *bhitər* or *bhitri*, door is *duər*. Terms like *pukhri* 'pond', *bande* 'id.' are of Aryan origin. Normally the kinship terms and terms for body parts are kept intact. So are the numerals. But here, the Aryan influence is so strong that all three areas are affected to some extent, giving way to foreign elements. Thus, wife is *bəhu*, nephew *bhəgnə,* niece *bhəgni*, *bhəi* is 'brother', hand is *hat,* one is *æk*. The present generation can only utter their numerals from one through five or six, then counting goes on in Aryan numerals. Terms for sentimental feelings like *maya* 'affection', *daya* 'pity', *laj* 'shame' are also borrowed. The first three ordinal numbers *pəhil/poilo* 'first', *dɔsar* 'second', and *tesar* 'third' are also borrowed from Indo-Aryan.

On the grammatical level, too, Aryan influence is noticeable. The distinction between masculine and feminine, though attested to in a few examples, is framed on the analogy of the Aryan system; the masculine noun ending in /-a/ and feminine in /-i/: thus *kala* 'deaf' masculine, and *kəli* 'id' feminine, *koŋka* 'mad' masc. *kuŋki* 'id', feminine, *koɽa* 'boy': *kuɽi* 'girl'. A good number of postpositions like *ləgit'* 'for', *sɔŋgɛ* 'alongwith', *upər* 'above', *bhitər/bhitri* 'in', *sathe* 'along with' are borrowed from either Bengali or Hindi. A good number of particles is also borrowed from the same source: *jodi/judi* 'if', *jɛmɔn* 'so that', *tahle* 'then/for that', *ar* 'and', *tɔkhɔn* 'then', *tɔbe* 'then' and so on. In conjugation also some roots like *cal* 'go', *bujh* 'understand', etc. are borrowed from the Aryan languages like Bengali, Hindi, or Oriya, but with modifications: */-ao/* or */-əu/* is added to the root resulting in *calao* or *bujhəu*. When these roots are conjugated the indigenous suffixes are used. TAM suffixes are never borrowed. Aryan nouns are also used as verbs but are so nativized that they fit well in the Santali system. In *bəhu-ad-e-a-ɲ* 'I gave him a wife', *bidə -ka-e-a-ko* 'they sent him off', the Santali suffixes are pitted against Aryan nouns. The non-native elements making their way into the vocabulary are readily nativized by giving a phonetic twist, and thus they are well incorporated in the language. In the non-native elements, following the native pattern the vowel */a/* is neutralized if it is preceded or followed by */i/* or */u/*: thus *bhitər* for *bhitɔr* or *bhetor* 'inside', *kəli* for *kali* 'goddess Kali', *kəmi* for *kam* 'work', *juən* for *juan* or *joan* 'youth', *bəhu* for *bahu* 'wife', *pəchim* for *pachim* 'west', *cəukidar* for *caukidar* 'village guard', *upər* for *upar*

'on the top, above', *sutəm* for *suta* 'thread', *məndir* for *mandir* 'temple', *ləḍu* for *laḍḍu* 'sweetmeat', *səɽi* for *saɽi* 'cloth', *rəni* for *rani* 'queen', *ʈhãi* for *ʈhãi* 'place', *utər* for *uttar* 'north', *rəi* for *rai* 'mustard', *ləgit'* for *lagi* 'for', *məmi* for *mami* 'maternal aunt', etc. Borrowed verb roots which are nativized by adding /*-ao*/ or /*-əu*/ like *paɽao* 'to fall', *calao* 'go', *jɔɽao* 'to link', *lagao* 'cultivate' the borrowed words or grammatical elements ending in a vowel also add a checked consonant at the end: thus – *ləgit'* for *lagi, ʈak'* for *ʈa* definite article.

1.6 Dialectal differences

Santali being scattered in different places is subject to differences in phonology, morphology, and lexicon. Considering that there has been no settled standard, R.N. Cust mentioned four dialects which, according to him, was not improbable.[15] In some reports the number of dialects is even more.[16] Although four or six is a bit ambitious, a line of demarcation is emerging gradually between the variety spoken in the districts of East and West Midnapur, Purulia and the southern portion of Bankura (comprising Khatra, Ranibandh, Raipur, Taldangra, and Simlapal blocks) in West Bengal, Balasore, Mayurbhanj, and Keonjhar in Orissa and that spoken in the northern portion of the district of Bankura (comprising Chatna and Saltora blocks), in Birbhum, Malda, Dinajpur north and south, Murshidabad, Cochbehar, and Jalpaiguri in West Bengal, Santal Parganas, Dumka, Singbhum, and Dhalbhum in Jharkhand, and Munghyr and Bhagalpur in Bihar. Campbell[17] mentioned two varieties like Northern and Southern, although he did not enter into the details. During our field trip we checked the varieties in detail and found two varieties in the areas named. Therefore Campbell's distinction of Southern versus Northern holds.[18] Differences between the two dialects – Northern (henceforth NS) and Southern (henceforth SS) are as follows.

1.6.1 Phonology

In the SS the pronunciation of [*ə*] is changing and gradually being replaced by [a], especially in the pronunciation of the younger generation, whereas in the NS it remains intact.

Loss of nasal and its compensation by nasalizing the preceding vowel is the rule of the SS, whereas in the NS the nasal remains intact.

Compare:

SS	NS	
põʈ	*ponḍ*	'white'
mãjhi	*maɲjhi*	'village headman'
ãʈia	*ənḍiə*	'male cow'
blhĩdar	*bhindər*	'to fall'
aḍoŋ	*anḍok'*	'to come out'
cãdo	*cando*	'sun'
nɔ̃ʈɛ	*nɔnḍɛ*	'here'
ɔ̃ʈɛ	*ɔnḍɛ*	'there'
hãʈɛ	*handɛ*	'there yonder'

Pronunciation of post-alveolar stop as flap is a common feature of the SS as opposed to the NS where it is pronounced as a stop.

Compare:

SS	NS	
põɽ	*ponɖ*	'white'
baɽa	*baɖa*	'to know'
huɽiŋ	*huɖiŋ*	'younger'
gɛɽɛ	*gɛɖɛ*	'swan'
bheɽa	*bheɖa*	'ram'

There is preference for nasals to checked consonants in case of genitive suffixes for inanimate, enclitic definitive and allative suffix /-sɛc'/ in the SS as opposed to the NS which preserves the checked ones.

Compare:

SS	NS	
/*-aŋ*/	/*-ak'*/	'genitive suffix for inanimate'
/*-rɛaŋ*/	/*-rɛak'*/	'genitive suffix for inanimate'
/*ɻaŋ*/	/*ɻak'*/	enclitic definitive
/*sɛn*/	/*sɛc'*/	'towards'

In the SS, sometimes there is a tendency, also shared by the younger generation of the NS to pronounce /e/ as [i] and /o/ as [u]. In certain examples like *abin* 'you two' (cf. *aben* in NS), *unku* 'they two' (cf. *onko* in NS), *ni* 'this' (cf. *ne* in NS) the pronunciation has been established. Among the elders and the enlightened, in the NS /*e*/ and /*o*/ are pronounced as half-close front and back vowels respectively.

1.6.2 Morphology

In morphology SS has certain features which the NS does not possess.

In the genitive case in the SS two suffixes are found when the governed noun is animate, one for the singular and the other for plural. /*-ic'*/ is used when the noun governed is in singular and /*-rɛn*/ when it is plural. In the NS /*-rɛn*/ is for both singular and plural.

Example:

SS		NS	
iɲ-ic' mɛrɔm(SG.)	'my goat'	*iɲ-rɛn mɛrɔm*	'my goat/ goats'(SG. and PL.)
iɲ-rɛn mɛrɔm	'my goats'		

The causative stem forming suffix is /*-hɔcɔ*/ in the SS and /*-oco*/ in the NS, although this is more phonological than morphological.

1.6.3 Vocabulary

In the lexicon SS and NS are somewhat different, initiated by borrowing from the neighbouring languages. The local borrowings in the two dialects are so high that sometimes one appears to be unintelligible to the other. In certain cases the usage

is also different. Terms like *ɛŋga* 'mother', *ɛra* 'wife', *hɛrɛl* 'husband', etc. are treated as vulgar in the SS and are almost obsolete, whereas in the NS the terms are regularly used without any pejorative sense. In the SS the corresponding terms for the relations are *ayo*, *bəhu,* and *jãvãy* are used for the kin relations respectively. Other examples are, for 'cloth' SS has *lugri* and NS has *kicric*', shirt is *dɔtɔ* in the SS and *jama* in the NS.

1.7 Orthography

Santali was first written down in Roman script by European missionaries. It was used then for translating the Bible into Santali, for writing grammars, and also for folk-tales and past history of the people. At that time many scripts, at least four, were being used for writing – like Devanagari, Bengali, Oriya and Roman. From the middle of the nineteenth century through the third quarter of the twentieth century the language was mostly written in Roman script. Even the people themselves used to write in Roman only for obvious reasons. Regional or Devanagari scripts came to be used much later. After Pandit Raghunath Murmu developed a script of their own, namely Ol Ciki, movements started for its recognition. In the 1970s many schools were started to impart training in Ol Ciki with initiative of a group of the Santals. A group of writers started writing in Ol Ciki. Magazines, journals, and newspapers began to be published. To the present date a consensus has not been reached as to which script is to be adopted.

A good many writings, like *Kherwal Bɔŋso Dhɔrɔm Pũthi* by Ramdas Tudu, *Ol Dɔhɔ Onɔrhẽ* and *Liʈə Godet'* by Ramchand Murmu, *Darega Dhon* and *Bidhu Chandan* by Raghunath Murmu, *Bhurka Ipil*, *Bidak Bela***,** etc. by Saradaprasad Kisku have been published in Bengali script. Even Ramdas Tudu and Raghunath Murmu who had developed scripts for the language published their own books in the language using Bengali script. A good number of books were also written in Devanagari and Roman scripts – Stephen H. Murmu's *Hɔʈ Bapla Puthi*, G.C. Tudu's *Chandmala* (a collection of poems) and *Bakhra* are written in Roman. Since 1965 The Santali Literary and Cultural Society has been rendering remarkable service for the development of the Santali language and literature by means of publishing Santali books and journals in Roman script. Narayan Soren, Balkishore Baskey, Bhagbat Murmu, Babulal Murmu, and Manikchand Hansda are well-known writers of Santali literature. Many of their writings are published in Devanagari script. However, Manikchand Hansda and S.D. Besra wrote in Roman script. Some of the most important Santali magazines are *Pera Hɔʈ* (in Roman) since 1922, *Marshal Tabon* (in Roman) since 1946, *Hɔʈ Sɔmbad* (in Devanagari) since 1947, *Pachim Bangla* (in Bengali script) since 1956, *Jug Sirijɔl* (in Roman) since 1971. Apart from these many other periodic journals are being published from West Bengal, Jharkhand, Bihar, Orissa, and Assam. *Tetre* from West Bengal and *Upel* from Assam need mention. *Tetre* is published in Bengali script and *Upel* in Roman. Two monthly magazines are published from Bangladesh – *Aboak' kurumuTureak' Kurai* and *GoDet'*, from Rongpur and Dhaka respectively. Both are in Roman script and serve as a link among those of Bangladesh, India, and Nepal.

Santali has recently been recognized by the Government of India as an official language, and it has been included in the Eighth Schedule of the Constitution of India. The Ol Ciki script has also been recognized by the Government of West Bengal as a medium of publication of Santali textbooks. It has already been included

in the curricula of the Universities. Whereas two of the Universities, Burdwan and Vidyasagar, are flexible in their approach (i.e. teaching, framing of question papers, and writing the same, will be in Roman, Bengali, or Ol Ciki until a particular script is accepted on consensus), the Vishva Bharati University imparts through Roman script. Some of the Universities of Bihar and Jharkhand have introduced Santali as a modern Indian language and literature in their syllabi, liberally accepting both Roman and Devanagari. It is yet to be seen which script is accepted finally, as a good percentage of population is in favour of regional scripts, Roman, or Devanagari. They are opposing introduction of Ol Ciki. Seminars and conferences are being organized in the government and on the organizational level to resolve the issue. But the debate continues.

2 PHONOLOGY

Santali is a Munda language with an elaborate system of vowels and consonants when compared with other languages of the same group (cf. Bhattacharya 1954, 1975, Stampe 1963, Matson 1964, Biligiri 1965, Zide 1965, Fernandez 1967, Rao 1982, Osada 1992, Nagaraja 1999, Ghosh 2003). There are eight vowel phonemes with nasalized counterparts (except /*e*/ and /*o*/). But the nasal ones are used rarely as compared with their oral counterparts. Length is not phonemic, although according to Bodding (1922, 1929) all vowels may be short or long. It has post-alveolar consonants and aspirated stops. It also has checked consonants occurring finally. Although most of the aspirates are found in loanwords, in some words of the native system the aspirates do occur but their origin is doubtful. In some cases aspirates originate through syllable reduction like *aji*+*hanhar* = *ajhnar* ('husband/wife's elder sister'). Although aspiration, in Pinnow's version (1959), was absent in Proto-Munda, it has spread into the phonemic system of Santali. The status of the checked consonants is clear – these are final allophones of the corresponding stop consonants except for post-alveolar, which has none. The checked plosives (labial, alveolar, palatal, and velar) involve a closure in the glottal cavity checking release of the air. A weak release is audible, however, if it is immediately followed by the corresponding unchecked consonant. Except post-alveolar all plosives have nasals in the phonemic level, and occurrence of the post-alveolar nasal is predictable. There is post-alveolar non-aspirated flap without having its aspirated counterpart, as mentioned in some literature (Neukom 2001:5).

2.1 Vowels

Table 2.2 lists Santali vowels.

TABLE 2.2: SANTALI VOWELS

	Front		Central		Back	
High	*i*	*ĩ*			*u*	*ũ*
Mid-high	*e*		*ə*	*ə̃*	*o*	
Mid-low	*ɛ*	*ɛ̃*			*ɔ*	*ɔ̃*
Low			*a*	*ã*		

2.1.1 Vocalic allophony

Note that /ə/ is found mostly in the environment preceded or followed by /i/or /u/: for example, *əḍi* 'very much', *əgu* 'bring', *niə* 'this very', *inə* 'that very', *gəi* 'cow', *dəi* 'elder sister', *gəḍi* 'cart', *ʈuər* 'orphan', where the occurrence is predictable. In some cases there is no such predictable environment, like *əd* 'take possession of', *əs* int. (bullock or buffalo), *bəd* 'high lying rice-field', *bəḍ* 'flood', *rəɽ* 'a tune', *ḍər* 'branch of tree', *səbrə* 'tasteless', *ʈəyə* 'leave alone', *dəblə* 'too broad', *dətrə* 'a certain plant', *gəhlə* 'low pitched', etc. In a number of examples like *ən* 'law' (cf. Bengali/ Hindi, *ain* 'id.'), *əd* 'origin'(cf. Bengali/Hindi, *adi* 'original'), *əg* 'fire' (cf. Sanskrit *agni*, Bengali *agun*, Hindi *ag* 'id.'), *əglə* 'forthcoming' (cf. Hindi *agla~agila* 'id.'), *ək* 'sugar-cane' (cf. Skt. *ikṣu* 'id.'), *ãs* 'scales of fish' (cf. Hindi *ãis* 'id.'), *rəs* 'heap' (cf. Bengali *raʃi* 'id.'), *rət* 'night' (cf. Hindi, *rait* 'night'), *rəskə* 'joy' (cf. Hindi, *rasika* 'id.'), *dəɽ* 'run' (cf. Hindi *dauɽ*, Bengali *dauɽ* 'id.'), *jət* 'caste' (cf. Bengali *jati* 'id.'), *pəc'* 'five' (cf. Skt *paɲca* 'id.'), *dəl* 'pulse' (cf. Hindi *ḍail* 'id.'), *ḍən* 'witch' (cf. Hindi *ḍain*, Bengali *ḍaini* 'id.'), *ghəʈ* 'fault, sin' (cf. Hindi *ghaʈi* 'id.'), *bhəgnə* 'nephew' (cf. Bengali *bhagina~bhagna* 'id.'), *bəs* 'age' (cf. Hindi *bais,* Bengali *bɔyes* 'id.'); although there is no predictable environment on the surface level, the occurrence of /ə/ can be justified by comparing it with the Indo-Aryan words where there is [*i*] or [*u*] in the neighbouring syllable. In some of the examples like *pəc'*, *bəs*, etc. /ə/ occurs because of the neighbouring palatal consonant or semivowel, [c'] and [y], respectively. Bodding (1922,1929) referred to at least three 'resultant' vowels: *ạ, ẹ,* and *ọ.* Besides *ə* (Bodding's *ạ*) no other variety, either *ẹ* or *ọ*, is attested to in our data.

2.1.2 Distribution chart

Distribution chart for the vowels is provided as follows:

Vowel	Initial	Medial	Final
/i/	+	+	+
/e/	+	+	+
/ɛ/	+	+	+
/a/	+	+	+
/ə/	+	+	+
/ɔ/	+	+	+
/o/	+	+	+
/u/	+	+	+

2.1.3 Contrast pairs

Minimal and subminimal pairs are provided for illustration.

Vowels:

/i/:/u/:

il	'feather'	*ul*	'mango'
biriŋ	'unpaid'	*burum*	'lie down'
si	'plough'	*su*	'hiss'

/i/:/e/:			
ir	'reap'	*er*	INTR.(call)
bir	'forest'	*ber*	'sun'
ji	'smell'	*je*	CONJ.part.
/e/:/ɛ/:			
egar	'undertake'	*ɛgɛr*	'scold'
hel	'time'	*hɛl*	'bind together'
aɽe	'edge'	*aɽɛ*	'vicinity'
/e/:/o/:			
eɽak'	'other'	*oɽak'*	'uncover'
jel	'deer'	*jol*	'fire'
je	INDEF.PART.	*jo*	'taste'
/ɛ/:/a/:			
ɛm	'give'	*am*	2 SG. PR.
mɛt'	'eye'	*mat'*	'bamboo'
baɽɛ	'please'	*baɽa*	'frequently'
/ɛ/:/ɔ/:			
ɛr	'sow'	*ɔr*	'pull'
hɛl	'bind together'	*hɔl*	'quickly'
sɛ	'or'	*sɔ*	'smell'
/a/:/ɔ/:			
ak'	'bow'	*ɔk'*	'to smoke'
jak'	'touch slightly'	*jɔk'*	'sweep'
hɔra	'away,up'	*hɔrɔ*	'tortoise'
/o/:/ɔ/:			
oco	causative suff.	*ɔcɔk'*	'remove'
hon	demonstrative	*hɔn*	'son'
lo	'draw water'	*lɔ*	'burn'
/a/:/ə/:			
acar	'pickles'	*əcur*	'turn'
bal	'burn a hole'	*bəl*	'influence of Bonga'
paera	'swim'	*pəurə*	'distilled liquor'
/o/:/u/:			
ot'	'mushroom'	*ut'*	'swallow'
kol	'send'	*kul*	'tiger'
ato	'village'	*ətu*	'flow'
/e/:/ə/:			
eto	'broken in'	*ətu*	'flow'
ber	'time'	*bəɽ*	'a certain tree'
hela	'cut down abundantly'	*hələ*	'bravo'
/o/:/ə/:			
oco	causative suff.	*əcu*	'set to do'
hola	'yesterday'	*hələ*	'bravo'
cirə	'a piece of land'	*ciro*	'the sun grass'
/a/:/ã/:			
ak'	'bow'	*ãk'*	'bellow'
jak'	'touch slightly'	*jãk'*	'keep in order'
aya	'fem.attendant'	*ãyã*	'true'

/i/:/ĩ/:			
miru	‘parrot’	*mĩrũ*	‘rimless’
/ɛ/:/ɛ̃/:			
hɛc’	‘come’	*hɛ̃c’*	‘heavy’
/ɔ/:/ɔ̃/:			
hɔ	INTERJ.	*hɔ̃*	‘also’
/ə/: /ə̃ /:			
kər	‘sore on the leg’	*kə̃r*	‘thistle’
/u /:/ũ /:			
ut’	‘swallow’	*ũt̪*	‘camel’
miru	‘parrot’	*mĩrũ*	‘rimless’

Bodding (1922) gives several values of each of the vowels, that is, each vowel may be ‘narrow’ or ‘wide’ or ‘mid-mixed’, in his own words, depending on the environment in which it occurs. So far as our data goes, such distinctions are non-phonemic. Moreover, Bodding’s (1922) identification of six ‘modified’ vowels like *a̤,e̤, e̤, i̤, o̤* and *ṳ* could not be identified in our data except /ə/, that is, Bodding’s *a̤* and in line with Pinnow (1959:35) they can be discarded as non-phonemic.

2.1.4 Length and nasalization

Vowel length is not phonemic. A vowel can be long or short depending on whether it occurs in an open (long) or closed syllable (short). Nasalization is phonemic. Nasalized forms of all the vowels are attested in our data in a limited number of examples. Because of the paucity of data contrastive pairs could not be given.

2.2 Diphthongs

Bodding (1922, 1929) lists three kinds of diphthongs – descending (after glide), ascending (before glide), and level (both vowels are equally strong). He arranged the following diphthongs into the above three categories. He also suggests most of the Santali diphthongs belong to the first category.

Descending:	*ae, ao, a̤i, a̤u, eo, e̤o, e̤i, e̤o, iu, oe, o̤e, o̤i, and ui*
Ascending:	*ea, ia̤, oa, ua̤, and ui*
Level:	*ea, ia̤, io, iu, oa, ua̤*

For the ascending dipthongs he had the words ‘may be counted as’, and for the level ones he noted that ‘the last class is not, however, always pronounced as one diphthong; they are frequently, some of them generally, dissolved into two distinct syllables with a euphonic semivowel between them’ (1922:6). In our data, descending diphthongs are the most common ones, although with differences from those of Bodding, with a few ascending diphthongs. Level ones are found nowhere in our data.

Descending:	*ae, ao, eo, iu, ui, əi, əu, ei, ɔe, ɛo, oe, and oi*
Ascending:	*ea, oa, uə, iə, and io*

Table 2.3 details Santali vowel combinations.

TABLE 2.3: SANTALI VOWEL COMBINATIONS

V^1/V^2	i	e	ɛ	a	ə	ɔ	o	u
i					iə*		io*	iu
e	ei			ea*			eo	
ɛ							ɛo	
a		ae					ao	
ə	əi							əu
ɔ		ɔe						
o	oi	oe		oa*				
u	ui					uə*		

Note that wherever there is /*i*/ or /*u*/ in the combination /*a*/ becomes [*ə*]. It is also to be noted that in Bodding's (1922, 1929) data, diphthongs like '*ei*' were not attested to. He had '*e̯i̯*', '*e̯o*', and '*o̯i*', all of which are not found in our data. We found '*oi*' instead. '*ui*' always behaves like a descending one as opposed to that in Bodding where it is both ascending and descending. The so-called level ones of Bodding are always found dissolved into two syllables. For most of the combinations, except /*iə, io, ea, uə*/, the coda is empty. Except these four all other combinations may be considered as a nucleus followed by glide. Examples:

iu	*jiu*	'spirit'	*iə*	*niə*	'this very'
				tiək'	'lead'
	nĩuri	'forest tree'		*ənḍiə*	'male'
				bəṛiə	'with difficulty'
				iəte	'because of'
ei	*meila*	'fair'	*io*	*tiok*	'to reach'
	meilaŋ	INT. 'come'	*ea*	*barea*	'two'
				dea	'back'
				reaṛ	'cold'
				sea	'to rot'
eo	*eoṛa*	'to wind thread round the spindle'	*oa*	*noa*	'this'
				toa	'milk'
	heoa	'to accustom'			
	leoha	'mix with a liquid'			
			uə	*ruə*	'fever'
ɛo	*hɛo*	'to carry on heap'		*ruəṛ*	'to return'
	nɛota	'to invite'		*ṭuər*	'orphan'
	lɛoṭhɛ	'stick to'			
ae	*sedae*	'in the past'			
	aema	'many'			
	beṛhae	'round about'			
	maejiu	'woman'			
	naeke	'priest'			
ao	*são*	'in association'			
	barhao	'to increase'			
	batao	'to obey'			
	benao	'to make'			

əi	*ləi*	'to tell'
	əikəu	'to feel'
	kəi	'sin'
	dəi	'elder sister'
əu	*əuri*	'soon'
	bəisəu	'to establish'
	hiləu	'to shake'
	biḍəu	'to test'
ɔe	*hɔedak'*	'thunder shower'
	bɔeha	'brother'
	kɔe	'to ask for'
	ɔkɔe	'who'
oi	*moiḍoŋ*	'bald'
	noi	'listen, my girl'
	oi	INT.('yes')
	poi	'retribution'
	ḍoi	'wooden spoon'
oe	*hoe*	'to be'
	hoeo	'to shave'
	poesa	'pice'
	podoe	'puff'
ui	*uihar*	'fond remembrance'
	rui	'cotton'
	turui	'six'
	muiguc'	'dirty'

2.3 Vowel harmony

So far as the harmonic sequence is concerned, there appear to be certain restrictions that need mention here:

(i) In the same stress group, if /*i*/ or /*u*/occurs, /*ə*/ but not /*a*/ will occur; *niə* 'this very', *inə* 'that very', *ruə* 'fever', *busək* 'give birth to', *butəl* 'power', *bidə* 'to dismiss'.

(ii) /*a*/ but not /*ə*/ co-occurs with /*e o ɛ ɔ*/; *aben* 'you' DL, *abon* 'we' INCL, *ale* 'we' EXCL, *alo* prohibitive PART, *nahɛl* 'plough', *tahɛn* 'to stay', *atɛn* 'to listen', *asɛn* 'to lead', *aŋjɔm* 'to hear', *sadɔm* 'horse', *adɔm* 'many', *asɔl* 'chief', *benao* 'to make', *boŋga* 'evil spirit', *peа* 'three', *dela* 'invite to come'.

(iii) If there is /*ɛ*/ or /*ɔ*/ in the first syllable of a stress unit having more than one syllable there must always be /*ɛ*/ or /*ɔ*/ in the following syllables; *ɛgɛr* 'to scold', *ɛhɔp'* 'to begin', *ɛtɔm* 'right', *ɛnɛc'* 'to dance, *gɛjɛr* 'to break, *gɛndrɛc'* 'a rag', *gɔʈɔ* 'to help', *gɔʈɛn* 'part.', although diphthongs like *ɔe* and *ɛe* are found in a limited number of examples.

(iv) Although sequence like *e-i* and *o-i* are found high vowels seldom co-occur with mid-high or mid-low vowels (exceptions being: *bedin* 'pagan', *begari* 'forced labour', *bepari* 'trader'. These are all loanwords.).

(v) /*e*/ and /*o*/ never co-occur with /*u*/ in the same stress unit.

2.4 Consonant inventory

Aspirates in Table 2.4 are given in parentheses, as they seem to be borrowed, not native. As the language has borrowed extensively from the neighbouring Indo-Aryan languages, two sub-systems had developed; I presume one to be native and the other borrowed. In the native subsystem the contrast between voiced and voiceless plosives, except for retroflex, is neutralized in final position where they are replaced by corresponding checked sounds. Within the borrowed sub-system voiced and voiceless plosives contrast in all three positions. The borrowed sub-system has aspirated plosives, both voiced and voiceless. The native sub-system seldom uses aspirates except for a few words like *dhiri* 'stone', *ajhnar* (*əji* + *hanhar*) 'wife's or husband's elder sister'. Ghosh (1994) did not mention the aspirated plosives, as the words' aspirates are mostly, if not all, of Indo-Aryan origin. Taking together both the native and borrowed words, Neukom (2001:5), however, posits aspirates including an aspirated flap /*r̥h*/ in the phonemic system of Santali. Moreover he has, though hesitatingly, identified checked [p' t' c' k'] as phonemes, which may not be so. The distribution of the consonants is shown in two charts, obviously with the idea of showing the differences between the two.

2.4.1 Consonants

Table 2.4 lists Santali consonants.

Note that the borrowed sub-system is different from the native sub-system in that it has aspirated plosives in all the five places of articulation. In addition, it has final contrast of voiceless and voiced unaspirated plosives, which is not there in the native subsystem. Comparative charts showing distribution of consonants as seen in Tables 2.5 and 2.6 will illustrate the point.

A closer scrutiny of the Tables reveals that

(i) Whereas, within the native sub-system, the contrast between voiced and voiceless unaspirated plosives, except alveolar in the final position, is neutralized, it exists in the borrowed sub-system. There are, however, a few instances (e.g. *oḍok* 'take out', *əbuk* 'wash', *ɲut* 'dark', *ɔt* 'earth') where non-checked plosives occur finally, but those are just too few to lead to a conclusion. Neukom (2001:5) considers, perhaps in a reserved way, checked [p' t' c' k'] as phonemes, but in our data they are actually allophones of /p t c k/.

(ii) In the native sub-system [p' t' c' k'] are specified when before voiceless plosives, and nasals, when they occur finally, optionally appear as voiced [b d j g] before vowels. [t'] is also realized as [t̰'] before [k].

TABLE 2.4: SANTALI CONSONANTS

	Bilabial	Alveolar	Post-alveolar	Palatal	Velar	Glottal
Plosives:vl	*p*	*t*	*ʈ*	*c*	*k*	
VL.ASP.	*(ph)*	*(th)*	*(ʈh)*	*(ch)*	*(kh)*	
VD.	*b*	*d*	*ɖ*	*j*	*g*	
VD.ASP.	*(bh)*	*(dh)*	*(ɖh)*	*(jh)*	*(gh)*	
Nasal	*m*	*n*		*ɲ*	*ŋ*	
Fricative		*s*				*h*
Trill		*r*				
Flap			*ɽ*			
Lateral		*l*				
Glide	*w*			*y*		

TABLE 2.5: NATIVE SUB-SYSTEM

	-V	V-V	V- #
/p/[19]	[p]	[p]	[p']/--#
/b/	[b]	[b]	—
/t/	[t]	[t]	[t']/--#
			[t̪']/--[k]
/d/	[d]	[d]	—
/ʈ/	[ʈ]	[ʈ]	[ʈ]
/ɖ/	[ɖ]	[ɖ]	[ɖ]
/k/	[k]	[k]	[k']/--#
/g/	[g]	[g]	—
/c/	[c]	[c]	[c']/--#
/j/	[j]	[j]	—
/m/	[m]	[m]	[m]
/n/	[n]	[n]	[n]
		[ɲ]/--hpA	
/ɲ/	[ɲ]	[ɲ]	[ɲ]
/ŋ/	—	[ŋ]	[ŋ]
/s/	[s]	[s]	—
/h/	[h]	[h]	[h][20]
/r/	[r]	[r]	[r]
/ɽ/	—	[ɽ]	[ɽ]
/l/	[l]	[l]	[l]
/w/	—	[w]	—
/y/	—	[y]	—

TABLE 2.6: BORROWED SUB-SYSTEM

	-V	V-V	V- / #
/p/	[p]	[p]	[p]
/ph/	[ph]	[ph]	[ph]
/b/	[b]	[b]	[b]
/bh/	[bh]	[bh]	[bh]
/t/	[t]	[t]	[t]
/th/	[th]	[th]	[th]
/d/	[d]	[d]	[d]
/dh/	[dh]	[dh]	[dh]
/c/	[c]	[c]	[c]
/ch/	[ch]	[ch]	[ch]
/j/	[j]	[j]	[j]
/jh/	[jh]	[jh]	[jh]
/k/	[k]	[k]	[k]
/kh/	[kh]	[kh]	[kh]
/g/	[g]	[g]	[g]
/gh/	[gh]	[gh]	[gh]
/m/	[m]	[m]	[m]
/n/	[n]	[n]	[n]
		[ɲ]/--hpA	—
/s/	[s]	[s]	[s]
/h/	[h]	[h]	[h]
/r/	[r]	[r]	[r]
/ɽ/	—	[ɽ]	[ɽ]
/l/	[l]	[l]	[l]
/w/	—	[w]	—
/y/	—	[y]	—

Examples:

(1) *sap'-ed-a-e* 'He is catching.' vs. *sab-ed-a-e*
(2) *sab-a-e* 'He catches/will catch.' vs. *sap'-a-e*
(3) *sap'-ket̪'-ko-a-e* 'He caught them.'

(iii) Aspirate plosives, both voiced and voiceless, are part of the borrowed sub-system. Examples like *dhiri* 'stone' and *ɖhumbək'* 'crumple', as given by Neukom (2001:5), are either due to contraction of a vowel (*dihiri>dhiri*) or are borrowed from Indo-Aryan through a process of nativization (*k'* added to *ɖhumbə*).

(iv) Velar nasal occurs medially (mostly with homorganic nasals) and only finally, never initially in the native sub-system, whereas in the borrowed sub-system it does not occur at all.

(v) Palatal nasals do not occur in the borrowed sub-system.

(vi) /ɽ/ in both sub-systems does not occur initially. Aspirated /ɽh/, postulated by Neukom (2001:5), without any minimal pair is not attested to in our data. Bodding (1932) also does not give any.

(vii) Glides do occur in both, but only medially. [v] is not attested in our data. Bodding (1936) mentions this as a labiodental open voiced sound.

2.4.1.1 Consonant phoneme oppositions

(i) Voiceless vs. Voiced:

/p/:/b/:			
pon	'four'	*-bon*	1PR PL INCL
ape	2PR PL	*aben*	2PR DL
/t/:/d/:			
tala	'middle'	*dal*	'thrash'
ato	'village'	*adɔ*	'then'
/c/:/j /:			
celɐn	'who'	*jelɐn*	'length'
/k/:/g/:			
kalɔʈ	'pullet'	*galɔʈ*	'agree upon'
aka	verbal suff.	*aga*	'exaggerated'
/ʈ/:/ɖ/:			
ʈalao	'disobey'	*ɖalao*	'make over'
suʈur	'crunch'	*suɖur*	'pour down'
laʈ	'take possession of'	*lɛɖ*	'excrement of horse'

(ii) Unaspirate vs. Aspirate (In these pairs borrowed words are marked with 'L'):

/p/:/ph/:			
padao	'break wind'	*phada*	'open'(L)
paʈak'	'split'	*pharak*	'distance'(L)
ape	2 pr.pl.	*aphɔr*	'sow seed'(L)
/b/:/bh/:			
bar	'two'	*bhar*	'load'
lab	'profit'		
/t/:/th/:			
tala	'middle'	*thapa*	'slap'(L)
tis	'when'	*thir*	'still'(L)
latar	'below'	*lathak'*	'humpy'(L)
kitəb	'book'(L)	*kɔthnɛ*	'feign not to know'
/d/:/dh/:			
dam	'value'(L)	*dhama*	'basket'(L)
diri	'delay'(L)	*dhiri*	'stone'
bada	'loan of seed'	*badha*	'obstruction'(L)
		godhʈa	'a felled sapling'(L)
/c/:/ch/:			
cal	'go'(L)	*chal*	'bark'(L)
bickom	'rather'	*bichnəu*	'disentangle'(L)
/j/:/jh/:			
jal	'lick'	*jhəl*	'long'(L)
ajmao	'test'	*ajhnar*	'elder sister'
/ʈ /:/ʈh/:			
ʈaka	'rupee'(L)	*ʈhək*	'deceive'(L)
kiʈli	'kettle'(L)	*kuʈhri*	'room'(L)
/ɖ/:/ɖh/:			
ɖaŋra	'bullock'	*ɖhaŋ*	'stick'(L)
canɖbɔl	'tail'	*jaɖhna*	'phlegmatic'

/k/:/kh/:

kan	verb copula	*khan*	'if'(L)
bakɽa	'perverse'	*bakhra*	'share'(L)

/g/:/gh/:

gan	'about'	*ghanʈa*	'bell'(L)
bagɽa	mixed'	*aghrao*	'extend'(L)

(iii) Alveolar vs. Post-alveolar:

/t/:/ʈ/:

tala	'half'	*ʈalao*	'transgress'
at'	'lose'	*ãʈ*	'greatly'

/d/:/ɖ/:

dale	'be heaped'	*ɖala*	'scales of balance'
ad	'half'(L)	*aɖ*	'cover'

/r/:/ɽ/:

ara	'saw'	*aɽa*	'sort, kind'
bar	'two'	*baɽ*	'pool'

(iv) Trill vs. Lateral:

/r/:/l/:

rəi	'mustered'(L)	*ləi*	'tell'
arɛ	'nine'	*ale*	1PR.PL.EXCL.
ber	'time'	*bel*	'spread a mat'

(v) Nasals:

/m/:/n/:

ma	optative particle	*na*	postposition
eman	'do such'	*enaŋ*	'a while ago'
dɔm	'delay'	*dɔn*	'leap'

/n/:/ɲ/:

naram	'soft'	*ɲaɽam*	'suck in'
im	'lever'	*iɲ*	1PR SG.

/n/:/ŋ/:

enec'	'until'	*eŋga*	'mother'
mun	'seer'	*muŋ*	'kind of pulse'

(vi) Fricatives:

/s/:/h/:

sakam	'leaf'	*hako*	'fish'
busup'	'straw'	*baha*	'flower'
as	'hope' (L)	*ah*	INTRJ of pleasure

(vii) Glides:

/w/:/y/:

dawa	'claim'(L)	*daya*	'mercy'(L)

2.5 Suprasegmental phenomena

Santali has no phonemic tones or registers. Bodding (1929:6) also denies the existence of tone in the language. But level, rising, or falling intonation is significant in the

sentence. Stress is always on the second syllable of a word unless it is monosyllabic, irrespective of whether it is an open or a closed syllable.

2.6 Syllable structure

Santali words have the syllable structure of $(C^1)V(C^2)(C^1)$, where C^1 stands for all kinds of plosives, nasals (except /ŋ/), trills, laterals, and fricatives. C^2 stands for homorganic nasals. V is a vowel or nasal. In addition there are certain general constraints.

(i) Words can begin with any vowel, diphthong, and consonant other than /ɽ/, /ŋ/, /w/, and /y/.

  (a) The nucleus may be a vowel or nasal. Both /ŋ/ and /n/ are attested in this function. Forms like /*mahnder*/ 'day before yesterday' [ma.hn.der], /*mahŋga*/ 'dear' [ma.hŋ.ga], /*mahnɖa*/ 'direction' [ma.hɳ.ɖa] have a nasal at the nucleus of the second syllable.
  (b) The opposition [±voice] is neutralized in the coda in the native sub-system.
  (c) No native word has either /s/ or /h/ in the coda. Where these are found they are invariably Indo-Aryan loans, e.g. *as* 'wish', *bes* 'good', *bah* 'excellent'.
  (d) Non-native words with CC at the onset simplify CC either by vowel insertion or by dropping a C; for example, *prabhu* > *purbhu* 'lord', *skul* > *iskul* 'school'. With CC at the coda the same process is employed; for example, bench > *benci*, forest > *phɔres*, bottle > *bɔtɔl*.

(ii) An instance of CC at the coda is invariably a combination of C^2 and C^1; for example, *sendra* 'hunt', *mɔŋɖ* 'tail', *mɔŋj* 'beautiful'.

(iii) CCC in the middle of a word is realized as CC.C; for example, *sendra* [send.ra] and C at the onset of the next syllable is invariably alveolar trill.

Santali has the types of syllables illustrated below:

V	*ɛ.hɔp'*	'begin'				
VV	*ae.ma*	'many'				
VC	*ac'*	'self'				
CVC	*sɛn*	'go'	*dal*	'beat'	*rɔɽ*	'speak'
CV	*ə.gu*	'bring'	*ma.hŋ.ga*	'dear'		
CVV	*ba.ɖae*	'know'				
CVCC	*canɖ.bɔl*	'tail'	*gɛnd.rɛc'*	'rag'		
VVC	*ead*	'memorize'(L)				
CVVC	*ɽuər*	'orphan'				

It is to be noted that syllables consisting of V, VV, or VVC, are rare in Santali. The language has a predilection for CV structure, which Bodding (1922,1929) has described as 'open syllable' and Neukom (2001) as 'light syllable', closely followed by CVC, Bodding's 'close' and Neukom's 'heavy' syllable.

In a frequency counting, out of 1,858 syllables (taken from Macphail (1983) distributed over 1,081 words), CV is 885 (i.e. 47.631%), CVC 641 (i.e. 34.499%), V 104 (i.e. 5.597%), VC 75 (i.e. 4.036%), VV 10 (i.e. 0.538%), and CVV 122 (i.e. 6.566%). Other types are negligible.

Santali also has predominantly dissyllabic word structure. In the same counting it was observed that 773 (i.e. 71.507%) are dissyllabic; 269 (i.e. 24.884%) monosyllabic, and

31 (i.e. 2.867%) are trisyllabic (most of which are borrowed). Due to the predominance of dissyllabic words, native words which were supposed to have been trisyllabic have been reduced to dissyllabic by dropping a vowel off of the second syllable; for example, *a* + *kiriŋ* > *əkriŋ* 'sell', *haɽam* + <*p*> = *hapɽam* < *hapaɽam* 'ancestors'.

So far as the syllable structure of the stems is concerned, the following sequences are observed:

CV.CV	*ma.re*	'ancient'
CV.CVV	*ma.nao*	'honour'
CV.CVC	*da.pal*	'fight'
CVC.CVC	*dal.paŋ*	'half-naked'
CVCC.CV	*ɖaŋg.ra*	'bullock'
V.CV	*ə.ɖi*	'very'
V.CVV	*e.koi*	'same'
V.CVC	*ɔ.pɔr*	'pull mutually'
VC.CV	*ɔk.tɛ*	'time'
VC.CVV	*ul.ʧəu*	'reverse'
VC.CVC	*ɔn.tər*	'mind'
VV.CV	*ae.ma*	'many'
VV.CVV	*əi.kəu*	'feel'

Of these sequences, CV.CVC is predominantly used. Tri- and more syllabic words are rare. In words with more than one syllable the second syllable gets the stress.

2.7 Morphophonology

(i) Checked consonants of the verb stems optionally become voiced before finite /-a/ and marker for the imperative mood.

(4) *dak'* 'rain' *dag-a.* 'It rains/will rain.'
also *dak'-a*

(5) *sap'* 'catch' *sab-a-m* 'You catch.'
also *sap'-a- m*

(6) *rəput'* 'break' *rəpud-a-e* 'He breaks/will break.'
also *rəput'-a-e*

(7) *pɛrɛc'* 'fill' *pɛrɛj-a-e* 'He fills/will fill.'
also *pɛrɛc'-a-e*

(ii) Checked consonants of the TAM suffixes in the active (*t'*) are obligatorily voiced before finite /*-a*/ and vowel-initial pronominal suffixes (*iɲ* 'I' and *-e* 3PR SG.).

(8) *jɔm-ked-a-e*
eat-PST:A-FIN-3SG:SUBJ
'He ate.'

(9) *dal-ked-e-a-ko*
beat-PST:A-3SG:OBJ-3PL:SUBJ
'They beat him.'

(10) *dal-keʈ'-ko-a-e*
beat-PST:a-3PL:OBJ-FIN-3SG:SUBJ
'He beat them.'

(11) *sap'-led-e-a-kin*
catch-PLUP-3SG:OBJ-FIN-3DL:SUBJ
'They two caught him.'

This way *et'* > *ed* in present imperfect, *ket'* > *ked* in past, *let'* > *led* in pluperfect.

(iii) Checked consonants in verb stems become voiced when followed by middle voice suffix /*-ok'*/. Thus,

(12) *hij-uk'-a-e* (<*hɛc'* 'come')
come-M-FIN-3SG:SUBJ
'He comes/will come.'

(13) *dej-ok'-mɛ* (<*dɛc'* 'climb')
climb-M-2SG:IMP
'Climb.'

(iv) /*e*/ of TAM suffixes /*-et'*/, /*-ket'*/ and /*let'*/ is raised when followed by 1PR SG. suffix. Thus,

(14) *tɔl-kid-iɲ-a-e*
tie-PST:A-1SG:OBJ-FIN-3SG:SUBJ
'He tied me.'

(v) Verb stems ending in a vowel (especially /o/) get an increment [*-n*] in the imperative.

(15) *teŋgon-mɛ* (<*teŋgo* 'stand')
stand-2SG:IMP
'Rise up.'

(16) *topon-mɛ* (<*topo* 'bathe')
bathe-2SG:IMP
'Take bath.'

3 MORPHOLOGY

3.1 Nominal morphology

The morphemes that take case affixes and postpositions and that show agreement in the verb are classed as nominals with subcategories of nouns and pronouns. Nouns and pronouns are grouped under the same class, as they have the same functional load. The major difference between the two is that whereas on the paradigmatic axis nouns show opposition in number, (±gender), and case, pronouns show opposition in person, number, and case.

3.1.1 Number

Santali has three grammatical numbers, both on nominals and predicates – singular (unmarked), dual (marked by *-kin*), and plural (marked by *-ko*). Number marking is obligatory in the case of animate nouns. Pronouns have separate forms for all the three numbers in the first and second person. The third person pronominal form takes the dual and plural suffixes to form their dual and plural forms. In the third person animate there is a marker /*-i*/; in the interrogative and the indefinite, /*-e*/ in the singular, but those are more gender markers (animate and inanimate).

seta-ø 'dog'	*seta-kin 'two dogs'*	*seta-ko 'dogs'*
dog-SG.	dog-DL.	dog-PL.

nu-i-ø 'this person'	*nu-kin* 'these two persons'	*no-ko* 'these persons'
this-ANIM.-SG.	this-ANIM.-DL.(ANIM.)	this-ANIM.-PL (ANIM.)
no-a-ø 'this thing'	*no-a-kin* 'these two thing'	*no-a-ko* 'these things'
this-INAN.-SG.	this-INAN.-DL.	this-INAN.-PL.
ɔkɔ-e-ø 'who'	*ɔkɔ-e-kin* 'who two'	*ɔkɔ-e-ko* 'who all'
INTER.-ANIM.-SG.	INTER.-ANIM.-DL.	INTER.-ANIM.-PL.
ok-a-ø 'which'	*ok-a-kin* 'which two'	*ok-a-ko* 'which all'
INTER.-INAN.-SG.	INTER.-INAN.-DL.	INTER.-INAN.-PL.
jãhã-e-ø 'anyone'	*jãhã-e-kin* 'any two persons'	*jãhã-e-ko* 'anybody'
INDEF.-ANIM.-SG.	INDEF.-ANIM.-DL.	INDEF.-ANIM.-PL.
jãhã-ø 'anything'	*jãhã-kin* 'any two things'	*jãhã-ko* 'any all'
INDEF.(INAN.)-SG.	INDEF.(INAN.)-DL.	INDEF.(INAN.)-PL.

Note that, in case of demonstratives the gender suffix for the animate in the singular merges with the number suffixes in dual and plural, but the inanimate marker remains intact. This may indicate that /*-i* / is the animate gender marker which is dropped when dual and plural suffixes are present. This may also indicate that /*-kin*/ and /*-ko*/ are originally dual and plural markers for the animate, and their application to inanimate nominals is rather secondary. In case of interrogatives and indefinites, however, the animate gender suffix is consistently present in singular, dual, and plural.

The dual forms of the first and second personal pronouns are also used to denote singularity when these are used among certain kin-relations. Parents-in-law and children-in-law use second person dual in addressing each other, and use the exclusive form of the first personal pronoun in talking to each other and referring to themselves, when just one person is meant. This forms a sub-system of honorific usage.

(17) *ceka-en-a-ben* *bəhu*
how-PST:M-FIN-2DL:SUBJ daughter
'How are you, daughter-in-law?'

(18) *ruə-k'-kan-a-liɲ*
fever-M-COP-FIN-1DL.EX
'I am getting fever.'

There is no honorific pronoun in Santali. But nowadays there is a tendency among the educated Santals to use the dual form of the second personal pronoun to address and talk to a respected, senior, or unfamiliar person(s).

The plural also functions as an expression of singularity among certain kin-relations. Co-parents-in-law in addressing each other and talking among themselves use the inclusive form of the first person when just one person is meant.

(19) *hɛnda ho sumdhi, cet'leka menak'-bon-a* 'O co-parent-in-law, how are you ?
(20) *ədi muskil-rɛ-bon paɽao-aka-n-a* 'I am in great trouble'

Sometimes /*-ko* / with the addition of /*ta-*/, hence /*-tako*/, functions as the plural suffix.

ʈuɖu 'a Santal sept' > *ʈuɖu-tako* 'men of the Tudu sept'
hɔpɔnera-ɲ 'my daughter' > *hɔpɔnera-ɲ-tako* 'my daughters'

In the absence of more examples it is difficult to say anything definite concerning this, that is, whether it is plural or has something to do with belongingness.

3.1.2 Case

Case markers attach directly to the bare nominal. Santali cases can be divided into following categories:

- Core (subject and object): unmarked.
- Peripheral (comitative, genitive, instrumental, sociative, allative, ablative, and locative): marked on the nominal.

Core cases are unmarked on the nominal but are marked in the verb in the form of incorporated pronouns (if subject or object is animate) as transitive subject, intransitive subject, and transitive object. In the case of verbs taking two objects the so-called indirect object (dative?) is marked in the verb with the applicative prefix /*a*-/. If there are two objects and both are animate, only the indirect is marked in the verb, that is, the indirect object is raised to object with the applicative /*a*-/. Genitive is also marked in the verb when it denotes inalienable possession, in which case the prefix /*t*-/ is attached to the applicative forms of the pronouns; otherwise it is marked in the noun phrase and functions as an attribute. Consider the following examples:

(21) *gidrə rak'-ed-a-e* (/*-e*/ is marked for the transitive SUBJ.*gidrə*)
child cry-IMPRF-FIN-3SG:SUBJ
'The child is crying.'

(22) *phulmoni sɛn-ɔk'-a-e* (/*-e*/ marked for intransitive SUBJ.)
Phulmoni go-m-fin-3sg:subj
'Phulmoni will go/goes.'

(23) *dal-ked-e-a-e* (/*-e-*/ marked for transitive OBJ.)
beat-PST:A-3SG:OBJ-FIN-3SG:SUBJ
'He beat him.'

(24) *ləi-a-ko-a-e* (/*-ako-*/ marked for transitive OBJ. with applicative /*a*/)
tell-APPL-3PL:OBJ-FIN-3SG:SUBJ
'He will tell them.'

(25) *uni dɔ ac'-ak' jaŋga rəput'-akat'-t-ae-a*
3SG TOP 3S-GEN leg break-PRF:A-POSS-3SGPOSS-FIN
'He has broken his leg.' (/*-tae-*/ marked for possessor/undergoer of the action)

Table 2.7 lists case markers.

Genitive. There are two sets of suffixes for genitive – one for animate the other for inanimate. The suffix for the animate /*-rɛn*/ is used when the governed noun is animate and /*-ak*'/ is used when the governed noun is inanimate. For the inanimate there are other suffixes too, such as /*-rɛak'*/, /*-rɛaŋ*/, /*-rɛnak*'/, and /*-rɛnaŋ*/. The suffixes marked as 'other' are normally attached to inanimate nouns and demonstratives. In one example like *hapɽam-ko-rɛak' katha* 'story of the ancestors',

TABLE 2.7: CASE MARKERS

Case	Marker	Example	Syntactic function
Core			
Nominative	*ø*	*bɔeha*	Transitive SUBJ Intransitive SUBJ Transitive OBJ
Peripheral			
Genitive	*-rɛn* (anim) *-ak',rɛak'*(inan)	*iɲ-rɛn bɔeha* 'my boy' *iɲ-ak' ti* 'my hand' *dare-rɛak' jɔ* 'fruit of tree'	possessor
Comitative	*-ʈhɛn/-ʈhɛc'*	*bɔeha-ʈhɛn*	goal,place
Instrumental-Locative	*-tɛ*	*ɖaŋ-tɛ* *iskul-tɛ*	instrument, cause, motion
Sociative	*são**	*bɔeha-são*	association
Allative	*-sɛn/-sɛc'*	*oʈak'-sɛc'*	direction
Ablative	*khɔn/khɔc'*	*dare-khɔn*	source, origin
Locative	*-rɛ*	*dak'-rɛ*	spatio-temporal location

Note
Peripheral cases are marked on the nominals.

/*-rɛak'*/ is found to be attached with animate noun. Examples showing occurrences of genitive suffixes:

iɲ-rɛn mɛrɔm	'my goat'	*am-rɛn hɔpɔn*	'thy son'
ac'-ak' jaŋga	'his leg'	*abo-ak' oʈak'*	'our house'
oʈak'-rɛak' duər	'door of the house'	*dare-rɛak' jɔ*	'fruit of the tree'

Here genitive suffixes are used to express possessive relationship or belongingness to somebody or something.

Sometimes genitive suffixes also mark the topic of discussion:

noa-rɛak'	*mit'ʈaŋ*	*kəhəni*	*ləi-ad-iɲ-a-e*
this-GEN	one-CLSSFR	story	tell-APPL:PST:A-1SG:OBJ-FIN-3SG:SUBJ

'he told me a story about this' – denoting something about or concerning.

So far as formation of /*-rɛn*/ and /*-rɛak'*/ is concerned, Neukom (2001:29) tries to derive these from locative suffix /*-rɛ*/ with the 'additional element *-n*' (in case of/ *-rɛn* /) and inanimate /*ak'*/. Ghosh (1994) tries to derive /*-rɛak'*/ in the same way, but he could not derive /*-rɛn*/ from the same locative, as the meaning of the final *-n* (in /*-rɛn*/) was not clear. Neukom, too, has the same problem as he submits 'which has no clearly definable meaning' (2001:29). L. Burrows (1915:17) called this type of genitive a locative-genitive, as this type of genitive often denotes a belongingness of something to a place.

Comitative. The suffix for this case is /*-ʈhɛn*/ (with the variant *-ʈhɛc'*). Bhattacharya (1975:148–149) assigned this suffix a dative role in saying that 'when the verb has two objects, if both the objects are animate, the indirect object which precedes the direct object takes a dative suffix (or postposition) whereas the direct object is in

the accusative case, i.e. it is represented by an object particle in the verb…'. His examples are:

maeju-ʈhen gidrə-ɲ em-ked-e-a
woman-DAT child-1SG:SUBJ give-PST:A-3SG:OBJ-FIN
'I gave the child to the woman.'

gidrə-ʈhen maeju-ɲ em-ked-e-a
child-DAT woman-1SG:SUBJ give-PST:A-3SG:OBJ-FIN
'I gave the woman to the child.'

In my data /*-ʈhɛn* / is not found in the dative role and if there are two animate objects of the verb, the indirect object with the applicative marker is always shown in the verb rather than the direct object. Neukom (2001: 29) assigns the suffix dative status. But the examples he has cited (except examples 10 and 11) appear more comitative than dative. Data containing /*-ʈhɛn* / in my field notes are more comitative than anything. Consider the following:

(26) *iɲ-ʈhɛn mena-k'-a*
1SG-with exist-M-FIN
'It is with me.'

(27) *ona alo-ʈhɛn-e seʈer-en-khan-ge uni biɲ dɔ*
that light-near-3SG:SUBJ reach-PST:M-if-FOC 3SG snake TOP
hətiər-tɛ-e gɔc'-en-a
weapon-INS-3SG:SUBJ kill-PST:M-FIN
'Having been reached near the light that snake was killed by the weapon.'

(28) *uni buɖhi-ʈhɛn cɔlɔ-bon*
that old woman-near go-1PL.INC
'Let us go to that old woman.'

(29) *iɲ am-ʈhɛn noa katha ləi-ləgit'-iɲ hɛc'-len-a*
1SG:SUBJ 2-near this word tell-for-1SG:SUBJ come-PLUP:M-F
'I had come to tell you this.'

(30) *uni ato məɲjhi-ʈhɛn sɔpɔhɔt'-e ɲam-a*
3SG village headman-with help-3SG:SUBJ get-FIN
'He will get help with the village headman.'

Instrumental-Locative. The suffix /*-tɛ*/ is used to indicate two functions – instrumental and locative. With the locative function it indicates the place (as well as the time), the destination (reached) as against the allative /*-sɛn* / which indicates 'movement towards, direction', etc.

(31) *oka-tɛ-m cal-ak'-kan-a*
where-LOC-2SG:SUBJ go-M-COP-FIN
'Where are you going?'

(32) *iskul-tɛ cal-ak'-mɛ*
school-LOC go-M-2SG:IMP
'Go to the school.'

(33) *nəi-tɛ-ɲ* *cal-ak'-a*
river-LOC-1SG:SUBJ go-M-FIN
'I go/shall go to the river.'

(34) *tin* *somɔe-tɛ-m* *cal-ak'-a*
when time-LOC-2SG:SUBJ go-M-FIN
'At what time will you go?'

Functions as instrumental are:
(i) Instrument:

(35) *ɖaŋ-tɛ* *tiok'-mɛ*
stick-by pull down-2SG:IMP
'Pull down by the stick.'

(36) *uni* *ʈuʈi-tɛ* *cɛ̃ɽɛ ʈuʈi-led-e-a-e*
3SG arrow-INS bird pierce-PLUP:A-3SG:OBJ-FIN-3SG:SUBJ
'He had pierced the bird with the arrow.'

(ii) Cause:

(37) *uni* *bɔtɔr-tɛ* *dəɽ-akad-a-e*
3SG fear-INS run away-PRF:A-3SG:SUBJ
'He has run away out of fear.'

(iii) Manner:

(38) *khusi-tɛ-bon* *rəskə- k'-kan-a*
pleasure-INS-1PL.INC enjoy-M-COP-FIN
'We are enjoying with pleasure.'

(iv) Through, by means of:

(39) *tinək'* *gɔnɔŋ-tɛ-m* *hatao-a*
how price-INS-2SG:SUBJ take-FIN
'At what price will you take?'

(40) *noa* *kapaʈ-tɛ* *ca-lak'-mɛ*
this door-INS go-M-2SG:IMP
'Go through this door.'

(41) *noa* *duər-tɛ* *cal-ak'-mɛ*
this corridor-INS go-M-2SG:IMP
'Go through this door.'

Sometimes /*-tɛ*/ is also found with the subject in passive construction:

(42) *iɲ-tɛ-ɲ* *mak'-akan-a*
1SG-INS-1SG:SUBJ cut-PRF-M-FIN
'It is cut by me.'

Sociative. Sociative is marked by a postposition *são* 'with, in association with'.

(43) *alɛ-são* *hij-uk'-mɛ*
1PL.EX-SOC come-M-2SG:IMP
'Come with us.'

(44) *pɔtɔb-são seleʈ hɔ̃-e əgu-i-a*
book-SOC slate too-3SG:SUBJ bring-y-FIN
'He will bring slate with books too.'

são is also found to occur with suffix /*-tɛ*/, thus *sãotɛ*:

(45) *hɔpɔn-são-tɛ bana eŋga hɔ̃-e gɔc'-led-e-a*
cub-SOC-INS bear-FEM. too-3SG:SUBJ kill-PLUP:A-3SG:OBJ-FIN
'He killed a she-bear along with her cub.'

Allative. The allative suffix /-sɛn / (variant /-sɛc'/) expresses direction towards destination.

(46) *gɔʈa-sɛn-gɛ dhar hena-k'-a*
all-ALL-FOC sharpness exist:M-FIN
'There is sharpness in all directions.'

bir-sɛn 'towards the forest' *buru-sɛn* 'towards the mountain'

Ablative. The postposition for the ablative is khɔn (with variant khɔc') and is used to express 'from, away from':

(47) *dare-khɔn ãʈgon-mɛ*
tree-ABL come down-2SG.:IMP
'Come down from the tree.'

(48) *səʈim-khɔn ɲɛl-ɔk'-kan-a*
roof-ABL see-M.-COP-FIN
'It is being seen from the roof.'

khɔn is also found to occur with inanimate suffix /*-ak'*/:

(49) *raja oʈak'-khɔn-ak'-e dəʈ-ked-a*
king house-ABL-INAN-3SG:SUBJ run away-PST:A-FIN
'The king fled away from the house.'

Locative. The suffix for the locative is /*-rɛ*/, indicating location in spatiotemporal axes. It specifies destination or place reached. So, if /*-tɛ* / indicates locative of motion, /*-rɛ*/ indicates locative of rest.

khaʈaŋ-rɛ 'in the ditch' *siŋgəʈ ɔktɔ-rɛ* 'at night fall'
ato-rɛ 'in the village' *ɲutət'-rɛ* 'in the dark' *bhitri-rɛ* 'inside'

3.1.3 Person

Santali distinguishes between alienable and inalienable possession. In the case of alienable possession, the possessor appears in the genitive before the head noun. In case of inalienable possession, instead of a genitive attribute the head noun itself is marked for the possessive relationship by any one of the suffixes for the three persons, namely *-ɲ* for first person, *-m* for second person, and *-t* for third person, irrespective of singular, dual or plural.

The system applies to kinship relational terms only. As in:

bɔkɔ-t	'his/their younger brother'	*bɔkɔ-ɲ*	'my/our younger brother'
apa-t	'his/their father'	*ɛŋga-t*	'his/their mother'
ɛŋga-m	'thy/your mother'	*kimin-mɛ*	'thy/your wife'
hɔpɔn-iɲ	'my/our son'		

The full form of the first person is *iɲ* instead of *-ɲ*, and the second person imperative form *mɛ* for *-m* is used when the term for the kinship ends in a consonant.

3.1.4 Definiteness

Santali has one marker *-tɛt'* for nouns to mark definiteness and another *-ɟak'* for pronouns. To mark unspecified objects Santali has none; the numeral for 'one' may be used to mark nondefinite referential entity. The definite marker is used to emphasize, especially identity.

(50) *gidrə-tɛt'-ko-e* *ɛm-at'-ko-a,* *ale dɔ baŋ*
child-DEF-3PL:OBJ-3SG:SUBJ give-APPL:PST:A-3PL:OBJ-F 1PL.EX TOP NEG
'He gave only to the children, not to us.'

(51) *ɖər-tɛt'* *əgu-i-mɛ,* *dare-tɛt'* *ikə-kə-k'- mɛ*
branch-DEF bring-y-2SG:IMP tree-DEF be-MOD-M-2SG:IMP
'Bring the branch, let the tree be.'

3.1.5 Nominal class/gender

Santali has two types of gender distinction, one grammatical and the other lexical. The grammatical gender distinguishes, on the one hand, between [±animate], which is a native system, and between [±male] with the inflections /*-a*/ and /*-i*/, respectively, which is borrowed from Indo-Aryan. The lexical gender is, in Bhattacharya's (1976:195) term, 'compounded sex-based gender', that is, sex-linked words are attached to the sexually indeterminate nouns to mark male or female.

Santali distinguishes between two classes of nouns – animate and inanimate.

From verb-roots and stems, derived demonstratives as well as lexical attributes, indefinite pronouns, and derived adverbials, animate nouns are derived with suffix /*-ic*'/ and inanimate nouns with /*-ak*'/.

rɔhɔy-ic'	'sower'	*rɔhɔy-ak'*	'that which is sown' < *rɔhɔy* 'to sow'
dadal-ic'	'man who beats'	*dadal-ak'*	'that which is beaten' < *dal* 'to beat'
nui-ic'	'his one'	*nui-ak'*	'his thing' < *nui* 'this one'
nɔtɛ-n-ic'	'one of this side'	*nɔtɛ-n-ak'*	'thing of this side'(< *nɔtɛn* 'belonging to this direction', 'in this direction'
ponɖ-ic'	'white one'	*ponɖ-ak'*	'white thing' < *ponɖ* 'white'
arak'-ic'	'red one'	*arak'-ak'*	'red thing' < *arak'* 'red'
ɛɟak'-ic'	'any one'	*ɛɟak'-ak'*	'anything' < *ɛɟak'* any'
jãhãn-ic'	'somebody'	*jãhãn-ak'*	'something' < *jãhã* 'any, some'
noŋka-n-ic'	'man like this'	*noŋka-n-ak'*	'thing like this' < *noŋka* 'in this manner'

The verb root with TAM suffixes can also take these suffixes, and then the entire verb construction becomes a noun, for example:

dal-ked-e-y-ic' 'one who struck him'

Besides living beings the following objects are also considered animate. Celestial bodies, that is, the sun, moon, and stars are considered animate. Consider the

following examples where the above are marked in the verb in the form of third personal pronominal marker:

(52) *sɨɲcãdo rakap'-kan-a-e*
sun rise -COP-FIN-3SG:SUBJ
'The sun is rising.'

(53) *ɲɨɲdəcãdo dubuc'-en-a-e*
moon set-PST:M-FIN-3SG:SUBJ
'The moon set.'

(54) *hola-rɛn ipil-ɨɲ ɲɛl-ket'-ko-a*
yesterday-GEN star-1SG:SUBJ see-PST:A-3PL:OBJ-FIN
'I saw the stars of yesterday.'

Spiritual beings are regarded as animate. Consider the following examples:

(55) *uni mit'-ʈaŋ kəli-boŋga benao-akad-e-a-e*
he one-CLSSFR Kali-idol make-PRF:A-3SG:OBJ-FIN-3SG:SUBJ
'He has made a Kali idol.'

The words 'puff-ball', 'ear-wax', 'complete shells of snails', 'thorn when pricked' are regarded as animate.

(56) *puʈkə-ko halaŋ-ket'-ko-a*
mushroom-3PL:SUBJ collect-PST:A-3PL:OBJ-FIN
'They collected mushrooms.'

(57) *ɛ̃rgɔt' mena-k'-ko-a*
ear-wax have-M-3PL:OBJ-FIN
'There are earwaxes.'

(58) *ɨɲ ɨɲ-rɛn jhinuk-ɨɲ rapak'-ket'-ko-a*
I 1S-GEN shells-1SG:SUBJ burn-PST:A-3PL:OBJ-FIN
'I burnt my shells.'

(59) *jənum-ɨɲ tot'-ked-e-a*
thorn-1SG:SUBJ extract-PST:A-3SG:OBJ-FIN
'I extracted the thorn.'

Dead human beings and animals are considered animate if the name is used or the person or animal referred to.

(60) *gɔc'hɔʈ-ko əgu-ked-e-a*
deadman-3PL:SUBJ bring-PST:A-3SG:OBJ-FIN
'They brought the dead man.'

(61) *gɔc' ənʈiə-ko topa-ked-e-a*
dead bull-3PL:SUBJ bury-PST:A-3SG:OBJ-FIN
'They buried the dead bull.'

The animate–inanimate distinction is also marked in the choice of genitive suffixes *-/-rɛn/* for animate and */-ak'/* with variants for inanimate.

Inflected gender. Sometimes the distinction between ±male is expressed morphologically. Certain nouns ending in *-a* form their feminine by replacing the final *-a* by *-i*. This is obviously an Indo-Aryan trait borrowed by Santali.

koʈa ‘boy’ : *kuʈi* ‘girl’;	*bhola* ‘dog’ : *bhuli* ‘bitch’
bheḍa ‘ram’ : *bhiḍi* ‘sheep’	*mama* ‘maternal uncle’ : *məmi* ‘maternal aunt’
kala ‘deaf’ : *kəli* ‘deaf.FEM.’	*koŋka* ‘foolish’ : *kuŋki* ‘foolish.FEM.’
koḍa ‘dumb’ : *kuḍi* ‘dumb.FEM.’	*caḍra* ‘bald headed’ : *cəḍri* ‘bald.FEM.’

Sex-based gender. This may be purely lexical where separate words for male and female are used to indicate different categories.

ənḍiə ‘ox’ : *gəi* ‘cow’	*bɔeha* ‘brother’: *misera* ‘sister’
jãwãy ‘husband’ : *bəhu* ‘wife’	*herɛl* ‘husband’ : *maejiu* ‘wife’
apat ‘father’ : *ɛŋgat* ‘mother’	*kaḍa* ‘he-buffalo’ : *bitkil* ‘she-buffalo’

Compounded sex-based gender. In this type the distinction between ±male is made clear by combining sex-based words like *ənḍiə*, *sanḍi*, *pēʈhar*, and *kuḍu* for male and *ɛŋga, bəchi*, and *pəʈhi* for female with sexually indeterminate words.

ənḍiə pusi ‘male cat’: *ɛŋga pusi* ‘cat’ FEM	*sənḍi sim* ‘cock’ : *ɛŋga sim* ‘hen’
pēʈhar mihũ ‘male calf’ : *bəchi mihũ* ‘do’ FEM	*kuḍu sukri* ‘boar’ : *pəʈhi sukri* ‘pig’

3.1.6 Pronouns (personal, interrogative, etc.)

The free-forms for the personal pronouns are given in Table 2.8. These pronouns may not be overtly marked in the utterance if the identity of the referent is not focused.

Note that there is an inclusive–exclusive difference in first person dual and plural. These forms are used in different social settings among certain kin-relations. They are not used indiscriminately. A more formal description would be that inclusive is [+speaker +hearer] where as exclusive is [+speaker –hearer], that is, the inclusive is marked by the presence of the addressee in the discourse while in the case of the exclusive it is not. There is also a distinction in the third person, that of anaphoric vs. unmarked.

Personal pronouns are not overtly marked if they are not focused.

(62) *khan ona aŋjɔm-katɛ goʈa bəd bɔehar-kin si-caba-ked-a.*
then that hear-CV whole high low land-2DL:SUBJ plough-finish-PST-FIN
‘Then having heard that they two ploughed up the whole land.’

(63) *am - gɛ ləi-mɛ tɔbe cikə-katɛ-ɲ gɔj-e-a*
thou- FOC tell-2SG:IMP then how-CV-1SG:SUBJ kill-3SG:OBJ-FIN
‘Then thou tell how shall I kill him.’

TABLE 2.8: PERSONAL PRONOUNS

Person	Singular	Dual	Plural		
		INCL	EXCL	INCL	EXCL
First	*iɲ*	*alaŋ*	*əliɲ*	*abo*	*alɛ*
Second	*am*		*aben*		*apɛ*
Third	*ac’*		*əkin*		*ako*
	uni		*unkin*		*onko*

In (62) the full form of the subject pronoun is not used as it is not focused, it is only marked in the predicate. In (63) the full form of 2SG. PR subject is used with the focus marker *gɛ* as it is focused.

Use of the dual or plural when one person is addressed. The first person exclusive dual *əliɲ* is used by a single speaker when s/he talks about or refers to him/herself to his/her father/mother-in-law, and first person plural inclusive is used by a single speaker when s/he talks about or refers to him/herself to his/her co-parent-in-law. The second person dual is used by parent-in-law in addressing son/daughter-in-law when just one person is meant.

(64) *hɛnda ho sumdhi cet'-leka mena-k'-bon-a*
o co-parent-in-law how-like be-M-1PL.INC-FIN
'O co-parent-in-law how are you (we)?'

(65) *əɖi muskil-rɛ-bon paɽao-akan-a*
very trouble-LOC-1PL:SUBJ fall-PRF:M-FIN
'I am in great trouble.'

(66) *ceka-en-a-ben bəhu*
how-PST:M-FIN-2DL:SUBJ daughter-in-law
'How are you, daughter-in-law?'

(67) *ruə-k'-kan-a-liɲ*
fever-M-COP-FIN.-1DL:SUBJ
'I am getting fever.'

Anaphoric vs. Unmarked. There are two sets of forms for the third person, one based on *ac'* and the other on demonstrative *uni* which is unmarked. The forms based on *ac'* are anaphoric while those based on the demonstrative *uni,* 'that person', are both anaphoric and deictic. *ac'* (with *əkin* and *ako* in dual and plural) is used for human referent only and refers back to the topic; otherwise *uni* (*unkin* and *onko* in dual and plural) is used.

(68) *ʈuər kuɽi gidrə. ac' baba bən-uk'-e-a*
orphan girl child. she father NEG-M-3SG:OBJ-FIN
'The orphan girl. She has no father.'

It can also be used in the same sentence when it is co-referential of the subject.

(69) *iɲ əgu gəi dɔ ac'-ak' goɽa-rɛ-y-e ader-akad-e-a*
1SG bring cow TOP he-GEN cowshed-LOC-y-3SG:SUBJ bring-PRF:A-3SG:OBJ-FIN
'He has brought the cow which I brought to his cowshed.'

Use of *uni* is anaphoric and deictic both. It should also be mentioned that frequency-wise use of *uni* is more common.

(70) *uni tɔkhɔn ɔnɖɛ-gɛ ərsi-tɛ ɲɛl-ɔk'-kan-tahɛ̃kan-a-e*
She then there-FOC mirror-LOC see-M-COP-PST-FIN-3SG:SUBJ
'Thereupon she was looking at the mirror.'

Forms of the interrogative pronouns are given in Table 2.9.

There are two types of interrogatives with ±animate distinction–referential and non-referential. The referential is used when the identity of the person or object is

TABLE 2.9: INTERROGATIVE PRONOUNS

	Animate	Inanimate
Referential	*ɔkɔe*	*oka*
Non-referential	*cele*	*cet'*

known to the addressor, and the non-referential when it is uncertain. Both the types form their dual and plural with *-kin* and *-ko*.

(71) *ɔkɔe- kan-a-m*
which person COP-FIN.-2SG:SUBJ
'Who are you?'

(72) *oka-ʈak' dɔ am-ak'*
which-DEF TOP thou-GEN
'Which (thing) is yours?'

(73) *cele kan-a-e*
who COP-FIN-3SG:SUBJ
'Who is there?'

(74) *cet'-t-am ɲutum*
what-POSS-2SG:SUBJ name
'What is thy name?'

Note that *ɔkɔe* can be used for human only, that is, it is + human, for non-human animate to be specified the definitive suffix *-ʈak'* is added.

(75) *oka-ʈak' dɔ am-ak'*
which-DEF TOP 2S-GEN
'Which one is yours?'

Other forms of Interrogative are

tis, 'when'; *tinək'* 'how much'; *cedak'* 'why'; *ceka* / *cikə* 'how'
tis is also found to occur with locative suffix *-rɛ*.

The most common indefinite root is *jãhã* 'any' which itself is used for the inanimate and it takes person marker *-e* for the animate. Both take *-kin* and *-ko* for dual and plural. Forms of the indefinite pronouns are given in Table 2.10.

(76) *jãhãe-gɛ met-a-e-mɛ*
anybody-FOC tell-APPL-3SG:OBJ-2SG:IMP
'Tell anybody.'

(77) *jãhã-tɛ cal-ak' ləi-oʈo-a-ɲ-mɛ*
any-LOC go-M tell-away-APPL-1SG:OBJ-2SG:IMP
'Tell me when thou go anywhere.'

(78) *jãhã disɔm-tɛ-bon dəɽ-a*
any country-LOC-1PL.EX:SUBJ go away-FIN
'We will go away to any country.'

TABLE 2.10: INDEFINITE PRONOUNS

Animate	Inanimate
jãhãe 'anyone'	*jãhã* 'anything'

Sometimes the inanimate form takes the definitive suffix *-ʈak'* to denote an inanimate object.

(79) *jãhã-ʈak'-gɛ* *hatao-mɛ*
any-DEF-FOC take-2SG:IMP
'Take any (of these).'

adɔm, 'some', is also found to be used as an indefinite for animates, and for inanimates it takes the inanimate suffix *-ak'* (i.e. *adɔmak'*), though more often it is used as an attribute rather than a pronoun. *ɛʈak',* 'another', is also used as an indefinite pronoun with *-ic'* for the animate and *-ak'* for the inanimate. Paradigms are listed in Table 2.11:

TABLE 2.11: OTHER INDEFINITES

	Animate	Inanimate
Singular	*adɔm* 'somebody'	*adɔmak'* 'something'
Dual	*adɔmkin*	*adɔmak'kin*
Plural	*adɔmko*	*adɔmak'ko*
Singular	*ɛʈak'ic'* another person'	*ɛʈak'ak'* 'another thing'
Dual	*ɛʈak'kin*	*ɛʈak'ak'kin*
Plural	*ɛʈak'ko*	*ɛʈak'ak'ko*

3.1.7 Demonstratives

Santali has two types of demonstratives with three distinctions – proximate, distal, and remote. While the proximate is in relation to the speaker, the distal and remote are in relation to the addressee. Two types of demonstratives are **simple**, represented by the roots *no* 'this', *on* 'that', and *han* 'yonder'; and **particularized**, represented by *ne* (also *ni*) 'just this', *en* 'just that', and *hen* 'just that yonder'. Both types have ±animate distinction and intensified forms as well. They may be represented in tabular form as shown in Table 2.12.

The intensified demonstratives are derived from both simple and particularized demonstratives for the proximate by infixing <*k'*>; thus *nuk'ui* 'this very one', *nɔk'ɔy* 'this very thing' from the simple type, and *nik'i* 'just this very one' from the particularized type. Note that only two forms, one for the animate and the other for the inanimate, are derived from the simple type without dual or plural, and only one from the particularized type for animate only. It is to be noted that when someone or something is to be specified, it is to be done with the proximate, and not with the distal or remote.

There are also two other types of demonstratives, one referring to what is seen and the other referring to what is heard, detailed in Table 2.13. In the case of the former there are two distinctions, one for the distant and the other for the remote. But in the case of the latter, there is only distant. While the demonstratives for what is seen are derived from *on* 'that' and *han* 'that yonder', the demonstrative of sound is derived

TABLE 2.12: DEMONSTRATIVES

Type I: Simple Demonstrative	Animate	Inanimate
Proximate	*nui* 'this person'	*noa* 'this thing'
DL	*nukin*	*noakin*
PL	*noko*	*noako*
Distal	*uni* 'that person'	*ona* 'that thing'
DL	*unkin*	*onakin*
PL	*onko*	*onako*
Remote	*həni* 'that person yonder'	*hana* 'that thing yonder'
DL	*hankin*	*hanakin*
PL	*hanko*	*hanako*
Type II: Particularized Demonstrative	**Animate**	**Inanimate**
Proximate	*nii* 'just this person'	*niə* 'just this thing
DL	*nikin*	*niəkin*
PL	*neko*	*niəko*
Distal	*ini* 'just that person'	*inə* 'just that thing'
DL	*inkin*	*inəkin*
PL	*enko*	*inəko*
Remote	*hini* 'just that one yonder'	*hinə* 'just that thing yonder'
DL	*hinkin*	*hinəkin*
PL	*henko*	*hinəko*

TABLE 2.13: DEMONSTRATIVES REFERRING TO SIGHT AND SOUND

	Demonstrative ref. to sight		Demonstrative ref. to sound
	Distal	Remote	Distal
Singular	*ɔnɛ* 'that seen over there'	*hanɛ* 'that seen over yonder'	*ɔtɛ* 'that heard over there'
Dual	*ɔnɛkin*	*hanɛkin*	*ɔtɛkin*
Plural	*ɔnɛko*	*hanɛko*	*ɔtɛko*

from *ɔt* (possibly *ot*). As both the forms end in *-ɛ,* it may be that it is a suffix pointing to some direction. The sets have no ±animate distinction.

3.1.8 Numerals

Numerals in Santali are generally found as quantifiers combined with classifiers. The actual number is denoted by the quantifiers used without classifiers in enumerating human beings in the indefinite, and with classifiers in enumerating human beings with definite and non-human animate and inanimate objects.

3.1.8.1 Cardinal

There are ten basic numerals in Santali, from 1 to 10, and further numerals are derived from these basic numerals. There are also two other numerals for 'twenty' and 'hundred', which are also used as bases for further numerals, but they are borrowed from Indo-Aryan.

Basic native cardinals				Borrowed	
mit'	'one'	*turui*	'six'	*isi*	'twenty'
bar	'two'	*eae*	'seven'	*sae*	'hundred'

pɛ	'three'	*irəl*	'eight'
pon	'four'	*arɛ*	'nine'
mɔ̃ɽɛ̃	'five'	*gɛl*	'ten'

There are two systems of counting, one based on 'ten', that is, decimal; and the other by scores, that is, vigesimal. From 'eleven' through 'nineteen' numerals are formed by adding numerals from 'one' to 'nine' to the base that is 'ten'. From 'twenty' onwards decades are formed by the process of multiplication and the numerals between the decades are the result of multiplication and addition. In multiplication the base is a multiplicand and that which multiplies is a multiplier. In addition the base is an augend and that which is added is an addend. The formation of the numerals may be given in a formulaic shape.

Numerals from 'eleven' to 'nineteen': AUGEND + ADDEND
gɛl + *mit'* = *gɛlmit'* 'eleven'
gɛl + *irəl* = *gɛlirəl* 'nineteen'

Numerals for decades : MULTIPLIER × MULTIPLICAND
bar × *gɛl* = *bargɛl* 'twenty'
pɛ × *gɛl* = *pɛgɛl* 'thirty'

Intermediate numerals: MULTIPLIER × MULTIPLICAND + ADDEND
bar × *gɛl* + *mit'* = 'twenty-one'
pon × *gɛl* + *eae* = 'forty-seven'

All the numerals are derived in this way until 'ninety nine'. For 'hundred' Santali does not have any native numeral. It is *sae* (cp. Skt. *ʃatam* 'hundred', Bangla *ʃɔ* 'id.'), borrowed from Indo-Aryan.

In the vigesimal system, that is, counting by scores, the same process operates, only the base for 'twenty' is *isi.* Even decades are multiples of 'twenty' and uneven decades are a combination of 'twenty' and 'ten'.

mit' + *isi* = *mit'isi*	'twenty'
bar + *isi* = *barisi*	'forty'
barisi + *gɛl* + *mit'*	'fifty-one'

3.1.8.2 Classifiers

There are three sets of classifiers in Santali.

Set I: *ʈɛn* (with variant *ʈɛc'*): This classifier is used with the numeral for 'one' and signifies non-living human beings, non-human living beings, non-human non-living objects and the inanimate. It is also used for human beings when specified.

mit'ʈɛn gɔc' hɔɽ	'one dead man'	*mit'ʈɛc' uric'*	'one bullock'
mit'ʈɛn rakkhos	'one man-eater'	*mit'ʈɛn hətiər*	'one weapon'
mit'ʈɛn hɔɽ	'one man'		

Set II: *ea*: This classifier is used with the numerals from 'two' to 'four' and for 'twenty', as well as the same classes of nouns as in Set I.

barea boŋga	'two ghosts'	*pɛa sim*	'three cocks'
ponea jinis	'four things'	*ponea gatɛ*	'four friends'

Set III: *gɔʈɛn* (with variant *gɔʈɛc'*): This classifier is used with the numerals from 'five' to 'ten' and with the distributive numerals. It is also used rarely with the numeral 'one'. Occurrences with 'one' are, however, found in songs and presumably demanded for metrical purpose. The classifier is used with the same classes of nouns as in Sets I and II.

mɔ̃ɽɛ̃-gɔʈɛn kaɖa	'five buffaloes'	*turui-gɔʈɛn boŋga*	'six ghosts'
eae-gɔʈɛn putul	'seven dolls'		

3.1.8.3 Distributive numeral

Distributive numerals in Santali are formed by reduplication of the initial consonant along with the vowel. Bodding (1929:60) has given examples of such formations, from 'one' through 'ten'. In my data only the first five numerals form their distributive counterpart in this process, and any other distributives are formed with *katɛ*. Thus:

mit' 'one' > *mimit'* 'one each'
bar 'two' > *babar* 'two each'
pɛ 'three' > *pɛpɛ* 'three each'
pon 'four' > *popon* 'four each'
mɔ̃ɽɛ̃ 'five' > *mɔ̃mɔ̃ɽ* 'five each'
turui-katɛ 'six each'

3.1.8.4 Inclusive numeral

Santali possesses three inclusive numerals – *banar,* 'both'; *pɛnɛ,* 'all three'; and *ponon,* 'all four' – derived by infixing <*n*> to the base numerals for 'two' to 'four'. They can be used as a subject or object of a sentence and as an attribute in endocentric attributive constructions.

(80) *banar-gɛ* *ʈuʈi-kin-mɛ*
both-EMPH pierce-3DL:OBJ-2SG:IMP
'Pierce the both.'

(81) *pɛnɛ* *əgu-akat'-ko-a-ɲ*
all three bring-PRF:A-3PL:OBJ-FIN-1SG:SUBJ
'I have brought all three.'

(82) *ponon* *kombɽo-ko* *gɔc'-en-a*
all four thief-3PL:SUBJ die-PST:M-FIN
'All four thieves died.'

3.1.8.5 Ordinal

Santali has three ordinal numbers – *pəhil* 'first', *dɔsar* 'second', and *tesar* 'third' – all borrowed from Indo-Aryan. Sometimes the ±animate suffixes *-ic'* and *-ak'* are also added. Each of them can occur as attribute or adverb.

Santali, being a neighbour of the eastern Indo-Aryan languages like Bangla, Hindi, Oriya, and Assamese, has direct contact with speakers of all these languages. Among other linguistic features, Santali also uses Indo-Aryan numerals. As a result, the first six numerals are used uniformly by all sections of society. The younger generation more often uses Indo-Aryan numerals from seven onwards while the older generation preserves the original numerals.

3.1.9 Adpositions

Santali has a large number of postpositions. The postpositions are added either to the bare nominals or to the number suffixes and the definitive marker. Some of them require the genitive case. Some are used after infinitives and others after a complete phrase – even after a sentence. Some postpositions are complex in the sense that they are composed of one postposition and a suffix. Some postpositions can take derivative suffixes for adjective and ±animate nouns. Here are some of the common postpositions:

lagat'/ləgit'. To indicate purpose or intention after all nominals and infinitives. It may also be combined with the suffix *-tɛ* without any change in meaning.

(83) *uni raj hɔpɔn-ləgit' dɔ bis ləḍu əgu-ad-e-a-e*
3SG king son-for TOP poison sweet bring-APPL:PST:A-3SG:OBJ-FIN-3SG:SUBJ
'He brought poisonous sweet for the prince.'

(84) *am-ləgit'-iɲ hɛc'-aka-n-a*
2SG-for-1SG:SUBJ come-PRF-M-FIN
'I have come for thou.'

(85) *uduk'-ləgit'-e calao-aka-n-a*
show-for-3SG:SUBJ go-PRF-M-FIN
'He has come for showing.'

modre. 'among'. Bangla word with *-rɛ*, i.e. *moddherɛ* is also found to occur in the same sense.

(86) *noko-modre kombɽo mena-e-a*
these-among thief have-3SG:OBJ-FIN.
'There is a thief among these persons.'

(87) *apɛ-moddherɛ jãhãe-gɛ lɔgɔn hij-uk'-mɛ*
2PL-among anyone-EMPH quickly come-M-2SG:IMP
'Any one of you come quickly.'

dhəbic'. Convey the sense 'till, until, up to'. Used after nominals indicating space and after infinitives indicating time.

(88) *cuɽa-dhəbic' ãɽgo-en-a-e*
peak-upto climb-PST:M-FIN-3SG:SUBJ
'He climbed up to the peak.'

(89) *uni əuri hij-uk'-dhəbic' okatɛ-hɔ̃ alo-m cala-k'-a*
3SG before come-M-until where-too PROHIB-2SG:SUBJ go-M-FIN.
'Do not go anywhere until he comes.'

bhitrirɛ. 'inside' with locative suffix *-rɛ*, hence 'within' too.

(90) *oɽak'-bhitrirɛ mena-e-a*
house-inside have-3SG:OBJ-FIN.
'He is there in the house.'

talarɛ. locative of tala 'middle' with locative suffix *-rɛ*.

(91) *oɽak'-talarɛ duɽup'-aka-n-a-e*
house-middle in sit-PRF-M-FIN-3SG:SUBJ
'He is sitting in the middle of the house.'

latarrɛ. locative of *latar* 'below' with suffix *-rɛ,* used only with noun.

(92) *dhiri-latarrɛ*
stone-under
'Under (a) stone.'

cetanrɛ. 'above,top', loc of *cetan* 'top',with suffix *-rɛ*.

(93) *ɖuŋri-cetanrɛ mit'-ʈɛn dare tahẽkan-a*
hill-under one-CLSSFR tree PST-FIN
'There was a tree under the hill.'

leka. Adjectival and adverbial postposition meaning 'like'. It is also found to occur with locative suffix of motion *-tɛ* in the sense of 'by any means'.

(94) *nui-leka hɔɽ kan-a-e*
DEM-like man be-FIN-3SG:SUBJ
'He is a man like this person.'

(95) *oka-lekatɛ hɔ̃ noa dɔ alo-m ɛm-ɔk'-a*
which-by means too this TOP PROHIB-2SG:SUBJ give-M-FIN
'Thou would not give it by any means.'

atɛ. 'taking alongwith'.

(96) *uni ləʈhi-atɛ taɽam-ed-a-e*
3SG:SUBJ stick-take along walk-PRS:A-FIN-3SG:SUBJ
'He is walking taking a stick along.'

hɔtɛc'tɛ. 'for, by, due to', indicating instrumentality.

(97) *dak' bɔdɛ-hɔtɛc'tɛ bə-ɲ topo-le-n-a*
water dirty- due to NEG-1SG:SUBJ bathe-PLUP-M-FIN
'I did not take bath due to dirty water.'

tuluc'. 'being with, association with', used with nominals and verb stems.

(98) *iɲ-tuluc'-e hɛc'-akan-a*
1s-along with-3SG:SUBJ come-PRF-M-FIN
'He has come with me.'

(99) *sɛn-ɔk'-tuluc'-e* *jəpit'-idi-ed-a*
walk-M –while-3SG:SUBJ sleep-simultaneously-IMPRF:A-FIN
'While walking he sleeps.'

katɛ. Gerundial postposition, meaning 'having done'. It is also used after nominals and adjectives in an adverbial sense.

(100) *ruəɽ- hɛc'-katɛ* *uni* *noa* *katha* *ləi-ked-a-e*
return -come-CV he this story tell-PST:A-FIN-3SG:SUBJ
'Having returned he told this story.'

(101) *phəsiara-katɛ* *hatao-ked-a-e*
deceitful-CV get-PST:A-FIN-3SG:SUBJ
'He got possession of it deceitfully.'

mɛntɛ. Lit. 'by saying'(*mɛn* 'say'+ *tɛ*), 'for the purpose of'.

(102) *ʈheŋga-mɛntɛ* *mak'-əgu-akad-a-e*
stick-for the purpose of cut-bring-PRF:A-FIN-3SG:SUBJ
'He cut and brought it for the purpose of making stick.'

iətɛ. Used after nominals and infinitives, with a sense of 'owing to, due to, on account of'.

(103) *əɖi* *hɔɽ-iətɛ* *ba-e* *sor-len-a*
many people-owing to NEG-3SG:SUBJ come close-PLUP:M-FIN
'He did not come closer on account of multitude of people.'

3.1.10 Derivation

There are three processes to derive nominals in Santali: prefixation, infixation, and suffixation. The processes are employed by verbs, lexical and derived adjectives, and nouns to derive the nominals.

3.1.10.1 Suffixation

The suffixes *-ic'* for animate and *-ak'* for inanimate may be described as nominalizers, too (Neukom 2001:57–58). They are used to form referential nominals from:

Reduplicated verb stems and verb roots

dadalic'	'one who engaged in beating'	< *dal*	'beat'
jɔmak'	'food'	< *jɔm*	'eat'
gɔc'ic'	'dead one'	< *gɔc'*	'die'
ɔlic'	'writer'	< *ɔl*	'write'
ɔlak'	'that which is written'		
rəhɔyak'	'that which is sown'	< *rəhɔy*	'sow'
ɛnɛc'ic'	'dance'	< *ɛnɛc'*	'dance'
kiriŋak'	'that which is bought'	<*kiriŋ*	'buy'

Simple and derived adjectives

ponɖic'	'white one',	*ponɖak'*	'white thing'	*ponɖ*	'white'
bogeic'	'good one',	*bogeak'*	'good thing'	*boge*	'good'

nɔtɛnic' 'one of this side', *nɔtɛnak'* 'thing of this side' <*nɔtɛn* 'belonging to this side' <*nɔtɛ* 'this side'
nɔnɖɛnic' 'one of this place', *nɔnɖɛnak'* 'thing of this place' < *nɔnɖɛn* 'belonging to this place' <*nɔnɖɛ* 'here'
noŋkanic' 'one like this one', *noŋkanak'* 'thing like this' <*noŋkan* 'like this' < *noŋka* 'in this manner'

Postposition

sãotɛnic' 'companion' < *sãotɛn* 'accompanying' < *são* + *tɛ*

Suffix

ɔltɛak' 'that with which is written(pen)' < *ɔl* 'write' *tɛ* 'instrumental suffix'

3.1.10.2 Infixation

The most productive process of nominal derivation in Santali is infixation. There are at least five infixes – <*tV*>, <*nV*>, <*mV*>, <*ɽV*>, and <*pV*> – of which <*tV*>, <*nV*> and <*pV*> are the most productive. The other two, <*mV*> and <*ɽV*> are used rarely. They are inserted into verbs, nouns, and lexical adjectives to derive nominals. The vowel of the infix (normally the first vowel if the root or stem is bisyllabic) is that of the root to which it is inserted, the exception being *ɛhɔp'* 'begin', which repeats the second vowel.

Examples:
<*tV*>: *bɔtɔr* 'fear', < *bɔr* 'to fear'; *rɔtɔk'* 'seam', < *rɔk'* 'sew'; *ɲutum* 'name', < *ɲum* 'to name'; *ɔtɔr* 'warp of a web', < *ɔr* 'draw/pull'; *ɛtɔhɔp'* 'beginning', < *ɛhɔp'* 'begin'; *gɔtɔrɔ* 'help', < *gɔrɔ* 'to help'; *ɔtɔmɔn* 'origin', < *ɔmɔn* 'germinate'; *satahɛt'* 'breath', < *sahɛt'* 'breathe'; *jeleɲ* 'long', < *jeteleɲ*, length'.

<*nV*>: *ɔnɔl* 'written piece', < *ɔl* 'write'; *jɔnɔk'* 'broom', < *jɔk'* 'sweep'; *banak'* 'hook', <*bak'* 'to hook'; *bɛnɛt'* 'stopper/lid' < *bɛt'* 'cover with palm'; *gɔnɔŋ* 'bride price', < *gɔŋ* 'give to marriage'; *tɔnɔl* 'knot', < *tɔl* 'bind'; *ranakap'* 'up, development', < *rakap'* 'rise'; *ɔɽ* 'begin' > *ɔnɔɽ* 'preface'; *ɔnɔsar* 'breadth', < *ɔsar* 'broad'.

<*pV*>: *hɔpɔn* 'children', < *hɔn* 'child'; *rapaj* 'king and his retinue', < *raj* 'king'; *məpəɲji* 'village chiefs', <*məɲjhi* 'village chief'; *kipisãɽ* 'zamindar and his retinue', < *kisãɽ* 'zamindar'. The infix <*pV*> is often employed to derive plural nouns.

<*mV*>: *lamak'* 'scraper', < *lak'* 'peal, scrap'; *cɛmɛt'* 'teaching aid', < *cɛt'* 'teach'; *sɛmlɛt'* 'association', < *sɛlɛt'* 'associate'; *hɔmɔn* 'nephew', < *hɔn* 'child'.

<*ɽV*>: *gɔɽɔm* 'grand old, namesake', < *gɔm* 'to name'; *cẽɽẽ* 'bird', < *cẽ* 'squeak'.

3.1.10.3 Prefixation

There is only one unproductive prefix in Santali, *ma-*, which serves as a nominalizer in converting verbs to nouns. It is very restricted in use, and only three verb roots

take this prefix for nominalization. In some cases it carries the meaning of result and in others that of the active agent.

marsal 'light', <*arsal* 'to light'; *marak'* 'peacock', <*rak'* ' cry'; *macɛt'* 'teacher', <*cɛt'* 'teach'.

3.1.11 Adjectives

There is little justification for positing a separate adjectival class except for a few derived adjectival words, which function as attributes in endocentric attributive constructions and as predicate complements. There are, of course, a handful of borrowed words that are adjectives at the source and thus may be treated as typical adjectives. The borrowed words show ± sex distinction, otherwise there is no gender distinction as such. Borrowed pairs like *khepa:khepi* 'mad', *kala:kəli* 'deaf', *koŋka:kuŋki* 'mad', *kãɽã:kəɽi* 'blind', *totra:tutri* 'stammering', *kõda:kũdi* 'dumb', and *phogra:phugri* 'toothless' are declined for gender with *-a* in the masculine and *-i* in the feminine.

-an is the suffix for deriving adjectives from nouns and demonstrative adverbs of location, manner, and direction: *daɽean koɽa* 'strong boy', <*daɽe* 'strength'; *kaɖawan hɔɽ* 'buffalo-having man', <*kaɖa* 'buffalo'; *dayawan kuɽi* 'kind girl', <*daya* 'kindness'; *nɔnɖɛn macɛt'* 'teacher of this place', <*nɔnɖɛ* 'here'; *nɔtɛn ɖaŋgra* 'bullock of this side' <*nɔtɛ* 'this side'; *noŋkan kəmi* 'work like this' <*noŋka* 'like this'.

The following are some examples of their use as a predicate compliment:

(104) *uni dɔ dayawan-kan-a-e*
3SG. TOP kind-COP-FIN-3SG:SUBJ
'He is kind.'

(105) *ona ɔnɔl dɔ noŋkan-gɛ-a*
that writing TOP like this-FOC-FIN
'That essay is like this.'

The word itself is not inflected to show degrees of comparison. To obtain comparative and superlative degrees, postpositions are used with the words with which something is to be compared. *khɔn* is used for the comparative and words of multitude with the same postposition for superlative.

(106) *hana dare noa dare-ko khɔn dɔ sẽɽa-gɛ-a*
that yonder tree this tree-PL-ABL TOP big-FOC-FIN
'That tree is bigger than this tree.'

3.1.12 Adverbials

There are three types of adverbs: simple, derivative, and words with locatives suffixes.

usəra 'quickly' as in (107) *usəra kəmi-mɛ*
quickly do-2SG:IMP
'Do quickly.'

dɔme 'very much' as in (108) *dɔme-ko rak'-ed-a*
very much-3PL:SUBJ cry-IMPRF:A-FIN
'They are crying very much.'

lɔgɔn 'hurriedly' as in (109) *lɔgɔn* *hij-uk'-mɛ*
hurriedly come-M-2SG:IMP
'Come hurriedly.'

eskar 'alone' as in (110) *eskar* *tahɛn* *baŋ* *boge-a*
alone stay NEG good-FIN
'Staying alone is not good.'

nahak' 'presently' as in (111) *nahak'-e* *hij-uk'-a*
presently-3SG:SUBJ come-M-FIN
'He will come presently.'

tis 'when' as in (112) *tis-em* *hij-uk'-a*
when-2SG:SUBJ come-M-FIN
'When will you come?'

Some other words are *nit* 'now', *cɔɽ* 'immediately', *laha* 'first', *un* 'then', *nɛtar* 'at present', *seday* 'in old days', *acka* 'suddenly', *pəhil* 'at first', *dinəm* 'daily', *enaŋ* 'then', *dhinəŋ* 'after a while'.

The suffixes *-ka, -ɖɛ, -tɛ* are added to the demonstrative and interrogative roots to derive adverbs denoting location, motion, and manner: *noŋka* 'in this manner'; *nɔtɛ* 'in this direction'; *nɔnɖɛ* 'here'; *okarɛ* 'where'; *okatɛ* 'to which side'; *ceka* 'how'.

Reduplicated words with or without *-tɛ* and words along with their echo counterparts are used as adverbs.

(113) *pəri pəri-tɛ* *kəmi-mɛ*
turn-INS do-2SG:IMP
'Do by turns.'

(114) *kəc' kəc'* *dak'-et'-tahẽkan-a-e*
little rain-IMPRF-COP:PST-FIN-3SG:SUBJ
'It was raining a little.'

(115) *hakopako* *hij-uk'-mɛ*
quickly come-M.-2SG:IMP
'Come quickly.'

(116) *sɛn-ɔk'* *sɛn-ɔk'-tɛ-e* *mɛn-ked- a*
go-M go-M-CV-3SG:SUBJ say-PST:A-FIN
'While going he spoke.'

3.2 Verbal morphology

The Santali verb is defined by the fact that it obligatorily takes tense, aspect and mood markers, markers for voice, pronominal arguments, marker for the finiteness of action, and sentential modality. The words that denote action, event, or condition are not those treated as verbs, but any word is treated as such provided it takes the above markers. The verb with the finite marker *-a* is treated as finite, and without this and the gerundial suffix is treated as non-finite.

Typical of the Munda languages and Santali is the fact that the verb may be any type of lexeme, any nominal with or without a genitive suffix, interrogatives, and indefinite stems, besides an exclusive class of verb roots. In the case of nominals,

interrogatives, and indefinites, there is always the need of a copula (light verb) to be used with it. Consider the following examples:

(117) *uni dɔ* *am-ak'-kan-a- e*
3SG TOP 2SG-GEN-COP-FIN.-3SG:SUBJ
'He is yours.'

(118) *ɔkɔe-kan-a-m*
who-COP-FIN-2SG:SUBJ
'Who are you?'

(119) *oka-rɛak'-kan-a*
which-GEN-COP-FIN.
'Where does it belong to?'

(120) *ɔjɔn-ad-e-a-ɲ*
medicine-APPL:PST:A-3SG:OBJ-FIN.-1SG:SUBJ
'I gave him medicine.'

(121) *mɔ̃i dɔ* *pargana-e* *kimin-ked-e-tiɲ-a*
daughter TOP Pargana-3SG:SUBJ daughter-in-law-PST:A-3SG:OBJ-POSS-1POSS-FIN
'A Pargana made my daughter his daughter-in-law.'

Note that in examples 117–121 a pronoun, interrogative, and noun are used in the predicate position with either a copula or a tense marker.

3.2.1 Subject

Santali is a nominative/accusative type of language in terms of verb agreement. The verb obligatorily agrees with the subject NP in terms of person and number. The subject is marked by the clitic pronominals in the verb phrase freely standing after the verb or with the word preceding it. The pronominals have the same shape as personal pronouns, except that the third person is marked by *-e, -kin,* and *-ko* in the singular, dual, and plural. When an animate noun stands as the subject NP, it agrees with the verb in the form of third person clitic pronouns. When the subject NP is a pronoun, it agrees with the verb by its clitic form. Table 2.14 gives an overview of subject markers on the verb:

TABLE 2.14: SANTALI AGREEMENT MARKERS

Person/Number	Singular	Dual		Plural	
		INCL	EXCL	INCL	EXCL
First	*-ɲ(iɲ)*	*-laŋ*	*-liɲ*	*-bon*	*-lɛ*
Second	*-m*		*-ben*		*-pɛ*
Third	*-e*		*-kin*		*-ko*

Note that the clitic form of the first personal pronoun is replaced by the full form *iɲ* when it is added to a consonant-ending word. The first person plural exclusive gets an increment *n* in its clitic form. The second person singular *-m* gets an increment *e* when added to a consonant-ending word, thus *em.* The third personal clitic forms

are actually number suffixes for the dual and plural. When a negative particle precedes the verb, the clitic pronominals are obligatorily marked in the particle. There is one marker, *-k',* for an inanimate subject used exclusively with the verbs *mena* 'to be' and *hena* 'to have'.

(122) *cɛt'* *jinis* *hena-k'-taben-a*
what thing exist-M-POSS-2nd dl.POSS-FIN
'What thing is there of you?'

(123) *uni-ak'* *oɽak'-rɛ* *mit'-ʈaŋ* *ʈhili* *mena-k'-a*
3s.-GEN house-LOC one-CLSSFR pitcher exist-M-FIN
'There is a pitcher in (his) house.'

It should be noted that the verbs *mena* and *hena* take the subject agreement in the position where the object normally comes, that is, not after a finite *a*, but before it.

Examples for illustrating positions of subject pronominal clitic:

(124) *khan-gɛ* *paʈoari dɔ-e* *dəɽ-ked-a*
then-FOC Patoari TOP-3SG:SUBJ run-PST:A-FIN

Also

khan-gɛ *paʈoari dɔ* *dəɽ-ked-a-e*
'Then Patoari ran away'

(125) *ba-e* *sɛn-len-a*
NEG-3SG:SUBJ GO-PLUP:M-FIN
'He had not gone.'

3.2.2 Object types

The marking of pronominal object plays an important role in Santali. The animate objects are marked in the verb in the form of infixed clitic pronominals. There are two types of objects in Santali: one direct, marked by the clitic pronominals in infixed form; and the other indirect, marked by an Applicative *a-*. In the simple present/future the Applicative is prefixed to the clitic pronominals, and with TAM it is prefixed to the TAM markers.

(126) *dal-a-ɲ-a-e*
strike-APPL-1SG:OBJ-FIN-3SG:SUBJ
'He strikes/will strike for me.'

(127) *dal-a-ɲ-kan-a-e* 'He is striking for me.'
-COP-

(128) *dal-ad-iɲ-a-e* 'He struck for me.'
-APPL.PST:A

(129) *dal-akawad-iɲ-a-e* 'He has struck for me.'
-APPL.PRF:A-

The clitic pronominals for the object stand after the TAM markers and before copula *kan* and *tahẽkan*. The clitic pronominals for the object are the same as the subject

forms, except for second person singular, which is <*me*>. Examples illustrating the position of direct object in the verb are shown in Table 2.15:

TABLE 2.15: DIRECT OBJECT PRONOMINAL INFIXES

	Singular		Dual		Plural
First	*dal-**iɲ**-a-e*	*dal-**laŋ**-a-e*	*dal-**liɲ**-a-e*	*dal-**bon**-a-e*	*dal-**lɛ**-a-e* 'He strikes me/us'
Second	*dal-**me**-a-e*		*dal-**ben**-a-e*		*dal-**pɛ**-a-e* 'He strikes thou/you'
Third	*dal-**e**-a-e*		*dal-**kin**-a-e*		*dal-**ko**-a-e* 'He strikes him/them'

Further examples:

(130) *dal-kid-**iɲ**-a-e*
strike-PST:A-1SG:OBJ-FIN-3SG:SUBJ
'He struck me.'

(131) *dal-led-**e**-a-ko*
strike-PLUP:A-3SG:OBJ-FIN-3PL:SUBJ
'They had struck him.'

(132) *dal-akat'-**lɛ**-a-e*
strike-PRF:A-1PL:OBJ-FIN-3SG:SUBJ
'He has struck us.'

(133) *dal-k-**iɲ**-a-e*
strike-OPT-1SG:OBJ-FIN -3SG:SUBJ
'He should strike me.'

(134) *dal-le-**m**-a-e*
strike-IRR-2SG:OBJ-FIN-3SG:SUBJ
'He would strike you.'

3.2.3 Tense

In Santali two copulas, *kan* and *tahẽkan*, express time dimension for the present and past, respectively. There is the concept of time for present, past, and future, and the aspect suffixes along with their own functions denote time dimension as well.

The copula *kan* denotes present time. It is quite different from other aspect-cum-tense suffixes. While aspect-cum-tense suffixes take an object after them, *kan* takes its object before it. It has, again, nothing to do with voice markers for active or middle; it can be used with both active and middle. It can also be independently conjugated, thus:

(135) *kan-a-ɲ* 'I am'
be-FIN-1SG:SUBJ

(136) *kan-a-m* 'Thou art'

(137) *kan-a-e* 'He is'

It is secondarily shifted to the non-completive determinative progressive in the middle and with first and third person singular object in the active.

(138) Active : *uni dal-iɲ-kan-a-e*
3sS beat-1SG:OBJ-COP-FIN-3SG:SUBJ
'He is beating me.'
Middle : *uni cal-ak'-kan-a-e*
3sS go-M-COP-FIN-3SG:SUBJ
'He is going.'

It is also used with the Iterative/Intensive stem to denote non-completive determinative progressive.

(139) *dadal-kan-a-e*
beat intensively-COP-FIN-3SG:SUBJ
'He is beating intensively.'

Corresponding to the *kan* copula for the past is *tahẽkan.* Like *kan*, it is independently conjugated.

(140) *tahẽkan-a-ɲ*	*tahẽkan-a-m*	*tahẽkan-a-e*
COP:PST-FIN-1SG:SUBJ	COP:PST-FIN-2SG:SUBJ	COP:PST-FIN-3SG:SUBJ
'I was'	'Thou were'	'He was'

Unlike *kan*, it never functions as an aspect marker. In combination with different aspect suffixes, it always denotes past action or state. It can combine with the copula *kan* as well to convey the sense of past progressive. The tense nuances expressed by this copula will be shown in the aspect-cum-tense network. Zero or the absence of any marker denotes the simple present/future.

3.2.4 Aspect

Santali employs a number of suffixes to denote different aspects. It is interesting to note that the suffixes employed to denote different states of action are also employed to denote time dimension, thus serving a dual role. Aspects in Santali may be divided broadly into non-completive and completive, each having two broad divisions – indeterminative and determinative. Non-completive indeterminative is indefinite, while the same determinative is progressive. Completive indeterminative is aorist while the same determinative is either resultative, stating an accomplished action with a result which is still present; or non-resultative, stating an action which happened long ago with a result which is no longer present or is an action to be taken up prior to some other action. In a tabular form the aspect framework of Santali may be represented as follows:

Non-completive		
Indeterminative:	Indefinite	*Ø*
Imperfective:	Progressive	
	Active	*-e*
	Active & Middle	*kan*
Completive		
Indeterminative:	Aorist	
	Active	*-ke*
	Middle	*-e*

Determinative	Resultative	*-aka*
	(accomplished with the result still present)	
	Non-resultative	*-le*
	(accomplished with the result not present)	

The non-resultative *-le* is also used to denote priorative action. All aspect suffixes except *Ø* and *kan* are added with active and middle voice markers when they are suffixed to the verb stems. Aspect suffixes with active and middle markers are shown in Table 2.16:

TABLE 2.16: ASPECT SUFFIXES WITH ACTIVE AND MIDDLE MARKERS

Non-completive		
Indeterminative	Active	*Ø*
	Middle	*-ok' Ø*
Imperfective progressive	Active	*-et'*
	Middle	*-ok' kan*
Completive		
Indeterminative aorist	Active	*-ket'*
	Middle	*-en*
Determinative resultative	Active	*-akat'*
	Middle	*-akan*
Determinative non-resultative	Active	*-let'*
	Midddle	*-len*
Determinative priorative	Active	*-le*
	Middle	*-len*

Note that determinative priorative active does not take any marker in the active. The aspect-cum-tense network can be enumerated as follows:

Non-past
Non-completive indeterminative as *simple present/future*:
Active *Ø*
Middle *-ok' Ø*
Non-completive determinative progressive as *present progressive*:
Active *-et'*
Middle *-ok' kan*
Past
Completive indeterminative aorist as *simple past*:
Active *-ket'*
Middle *-en*
With Applicative[21] Active *-at'*
Middle *-an*
With Benefactive[22] Active *-kat'*
Completive determinative resultative as *perfect*:
Active *-akat'*
Middle *-akan*