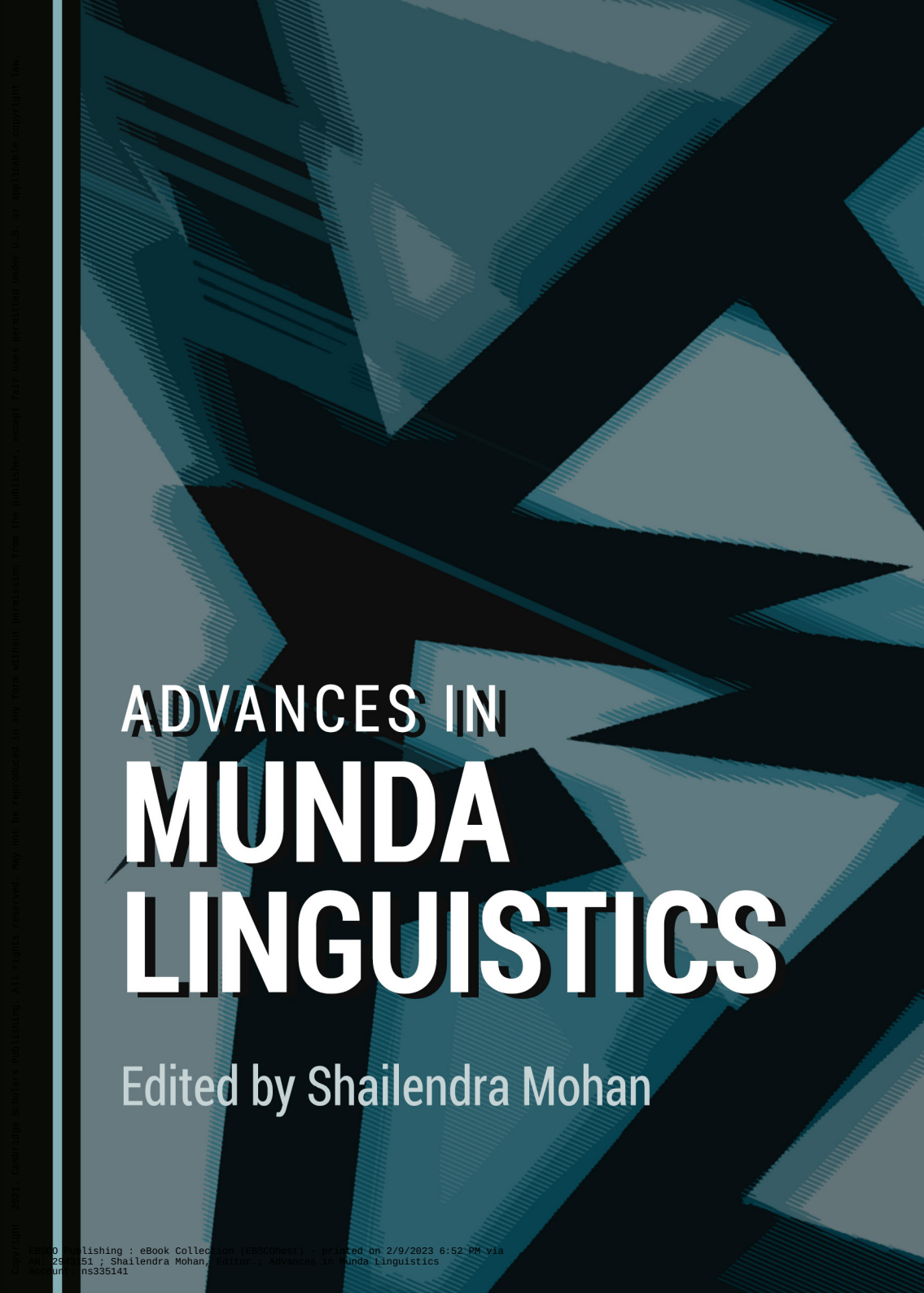

Advances in Munda Linguistics

Advances in Munda Linguistics

Edited by

Shailendra Mohan

Cambridge Scholars Publishing

Advances in Munda Linguistics

Edited by Shailendra Mohan

This book first published 2021

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

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

ISBN (10): 1-5275-6829-6
ISBN (13): 978-1-5275-6829-7

DEDICATED TO GREAT PIONEERS
IN MUNDA LINGUISTICS
AND TO MOTI, ARYENDRA, AND MY PARENTS

TABLE OF CONTENTS

List of Illustrations

LIST OF TABLES

ACKNOWLEDGMENTS

It gives me great pleasure to place before the scholar this edited volume containing the proceedings of the International Seminar on Munda Linguistics organized by the Department of Linguistics, Deccan College Post Graduate and Research Institute, Pune in collaboration with the Indian Council of Social Science Research (ICSSR), New Delhi and Central Institute of Indian Languages, Mysore.

I want to thank all the contributors for accepting my request and contributing their papers for this volume. I hope this volume would be useful to teachers, students, and researchers of linguistics, especially those working on Munda Linguistics.

I want to take this opportunity to express my sense of gratitude to the Indian Council of Social Science Research (ICSSR), New Delhi for its financial support to organise this seminar. I also thank ICSSR, New Delhi for the Joint Research project on 'Establishment of International Munda Studies Network' through ICSSR (India)-JSPS (Japan) Joint Research Programme in the field of Social Sciences. I also thank the Central Institute of Indian languages, Mysore for providing financial support.

I am highly obliged to my University, Deccan College Post Graduate and Research Institute, Pune, India, for allowing me to organise this seminar. I am beheld to the management committee of Deccan College, University's Vice-Chancellor Professor V.S. Shinde, Professor Jayashree Sathe (Pro-Vice-Chancellor), and C.V. Joshi (Registrar) for their encouragement and cooperation in organising the seminar.

My colleagues Sonal Kulkarni-Joshi, Rahul Mhaiskar, Shubhangi Kardile, Megna Carvalho, and Khusboo Parghi have helped me in one way or another in organising the seminar. I also thank Sambhaji Jadhav, Satish Bangar, Sanjay Hargude, and Mandar for their help during the seminar. I am thankful to all of them. Special thanks to Sanjay Hargude for harmonising the formatting of papers included in this volume.

My family-wife Moti, son Aryendra and my parents-have have been of great help during the seminar and also during the editing part of this volume.

Our special thanks are due to Cambridge Scholars Publishing, United Kingdom, and its anonymous reviewers for the support received during the publication process. Special thanks to Adam Rummens, Victoria Carruthers,

and Sophie Edminson for their patience and encouragement while the volume was in preparation. Last but not the least; I thank Prof. Claus Peter Zoller for his valuable suggestions.

Thank you all.

Shailendra Mohan

FOREWORD

Under the re-assuring blanket called the “Indian Linguistic Area”, the Munda languages harbour a wealth of unique phenomena, which will surely appear to anyone who cares to look into their particulars.

In the past, there have been abundant descriptions; grammars, dictionaries, even encyclopedias dedicated to the main Munda languages, e.g. Santali, Mundari, Sora, and also careful studies of the smaller ones. But the flow of published information on the Munda branch of Austroasiatic languages seems to have abated, compared to what it used to be. It is not that the languages are moribund or forgotten, by any means large amounts of data are being collected and preserved out of the public eye, in India and elsewhere. But the publishing of books, with their secure and easily accessible form, seems, unfortunately, to have hit a dry patch, at least for the moment.

This makes the publication of the current volume of Munda Linguistics a valuable sort of re-awakening. The expected linguistic data is there, and it occupies its rightful place, and then some of the more pointed and innovative fields of study in Linguistics are also well-represented in all their variety: experimental phonetics, semantic descriptions, even historical claims, and the more theoretically inclined.

Not only should that, but the variety of origins of the authors also be noticed: from India, as is legitimate, and from the West as is traditional, but also from the East, a welcome novelty. Among the Indian authors, one must also remark the discreet presence of some speakers of Munda languages; this is very valuable as they offer not only native knowledge in all its vast extent and subtlety of detail but also, and just as importantly, because of their own mastery of scientific argumentation.

And then, this is a tangible, printed book that readers can cherish and annotate at will, not an electronically erasable sequence of zero’s and one’s; a BOOK, with pages that can be turned back and forth in many ways with great ease, and with a hard, or a pleasing cover.

Gérard Diffloth
Siem Reap, Feb.2021

EDITOR'S PREFACE

The contribution in this book draws upon the empirical richness of the Munda languages spoken in India. This collection of papers makes an important contribution in terms of analysing and demonstrating key issues such as Proto-Munda reconstruction, migration of Munda language speakers, and synchronic linguistic issues in Munda languages spoken in India. This book would potentially be the first volume on Munda language and linguistics in which several indigenous Munda language speakers would contribute to the scholarship. The contributions will reflect the diverse range of scholarship on Munda languages, which combine empirical and theoretical discussion on issues related to Munda languages.

The first article "On the Role of Areal and Genetic Factors in the Development of the Word-Structure and Morphosyntax of the Munda languages" is from Gregory D.S. Anderson, who begins with a brief assessment of the current state of Munda language classification in India and contributes to the understanding of the role of contact vs. inheritance in the historical analysis of the Munda languages in India. He argues that there are features of Munda which might reflect pre-Aryan-Dravidian contact strata in the history of Munda, or otherwise lacks analogues in other Austroasiatic groups and in Indo-Aryan and Dravidian languages, but are reminiscent of features found in the periphery of South Asia now, in Munda, Tibeto-Burman (Trans-Himalayan), and in the isolate language Burushaski, which might reflect an earlier pre-Aryan-Dravidian profile of South Asia attested in Munda as well. The same contact strata also contributed to archaic retentions, contact-driven restructuring, and internal developments all separately manifested in Munda Grammar.

The second article in this volume "Infixation in Munda and its Austroasiatic Legacy" by Arun Ghosh presents the comparative study of infixation–a class-changing device in Munda and other Austroasiatic languages. According to him, infixation is an essential morphological process operated in both nominal and verbal derivations in Munda. In adjectival constructions too, infixation sometimes plays an important role. According to him, the infixation processes can be grouped into four parts namely, one, those which have developed in the individual languages like Santali, Korku, Sora, and Remo; two, there are some infixes like <-p-> which are shared by the major languages of the north Munda like Santali,

Mundari, Ho, and Korku; three, there are two infixes like <-t-> and <-m-> shared by some (not all) languages of both branches, north and south, like Santali, Mundari, Ho, Gutob, and Sora; and, four, there is one infix, <-n->, which is shared by all the languages of Munda and more than ten languages of the Mon-Khmer group. According to him, it may be assumed that the infixes developed in the individual languages well after the individual languages grew. <-p-> which is commonly shared among the North Munda languages may be postulated in the proto-North Munda. Similarly, *<-t-> may be postulated in the proto-Munda stage as some languages of both the groups share it. <-m-> as shared by some languages of both North and South Munda and most importantly by Khmer may be postulated in the proto stage of both the groups. The most common infix in both Munda and Mon-Khmer is <-n-> which with utmost certainty can be proved to be belonging to the proto-stage of both the groups.

The next paper "Agreement Reversal in Munda Languages: An Interplay of Functional/ Thematic and Syntactic Criteria" is authored by K.V. Subbarao, and Martin Everaert. In this paper, Subbarao and Everaert discuss the agreement reversal in North Munda languages (Santali, Ho, and Mundari) in which the Subject Agreement Marker (SAM) in oblique object constructions occurs not in its own canonical position, but in the position earmarked for an object, despite the fact that the predicate may be [-transitive]. They have argued that such a reversal takes place, not due to the syntactic principles governing the agreement, but due to thematic/functional criteria because of which the nature of the predicate in a non-nominative subject construction triggers such reversal and they conclude that agreement in the North Munda languages can be accounted for by invoking syntactic as well as thematic/functional criteria.

The fourth article "Performance in Elicitation: Methodological Considerations in the Study of Mundari Expressives" is authored by Nathan Badenoch, Nishant Choksi, Toshiki Osada, and Madhu Purti. In this paper, the authors present the methodological aspects of eliciting expressives in the Mundari language. According to the authors, expressives as a linguistic device in languages has not received attention because the semantics of expressives are complex as they deal with multi-sensory depictions that are situated in a speaker's direct experience. The authors conclude that the elicitation of expressives is notoriously difficult, but using a combination of story-telling and speaker enactments, a better idea of the semantic domains and pragmatic nuances of expressives can be obtained. This understanding can provide a deeper view of the grammar of the language but can also provide new perspectives on how language is embedded in cultural practices and world-views.

Masato Kobayashi presents the next paper on "The Past Suffixes of Hill Korwa". In the paper, the author presents the verb morphology of Korwa and concludes that the verb morphology of Korwa looks unique. Due to the sound changes and grammaticalization, and through a closer comparison with other Kherwarian verbs, it can be said that Korwa has a set of past suffixes similar to Ho, even though their distribution is more limited. Korwa verb morphology is characterized by the loss of the transitivity contrast by-*ḍ*/-*n*. Instead, -*ḍ* and -*n* are incorporated in tense and aspect suffixes in Korwa. The transitive-intransitive contrast is reduced, and only the originally transitive marker -*ḍ*/-*r* occurs in limited contexts.

In a joint paper, Bikram Jora and Gregory D.S. Anderson present the "Typologically Quirky Characteristics of Past and Perfective Forms in Kherwarian". According to the authors, Kherwarian Munda languages show a range of quirky features in their past or perfective forms. In past copula formations, in either existential/locational or possessive functions and constructions, the animate participant or referent is encoded as subjects in the past but as objects in the present, in negative and positive forms alike. While most Kherwarian languages show elaborate morphology in virtually all verbs, polyvalent negative aorist forms in Birhor show a subtractive pattern where all the morphology on the verb itself is suppressed, only subject clitics that appear on the negative scope operator before the bare verb stem are used. They also observe that in various Kherwarian languages a positive vs. negative opposition is emerging in the use of TAM markers in positive and negative constructions, such that the now general negative past/perfective marker *le*- may lose the anterior or pluperfect meaning it typically conveys in positive formations. Some of these features have analogues in other Munda languages like Sora or Korku and thus likely reflect retentions of earlier features.

The seventh article "Phrasal Affixes as Clitics in the Munda Languages" is written by Anish Koshy. In this paper, the author presents examples from different Munda languages of attachment of different bound elements that must be treated as phrasal affixation, that is, as clitics. The paper argues that a large number of clitics in the Munda languages are not just the agreement clitics that choose their hosts indiscriminately, as in Mundari and Santali. In many Kherwarian languages, like Mahali, Karmali, Turi, and Bhumij, the subject enclitics may appear only at the end of the verb complex yet are to be analysed as clitics due to their status as phrasal affixes.

The eighth chapter "Phonetic Comparison of Orissa Sora and Assam Sora" is jointly authored by Luke Horo and Priyankoo Sarmah. In this

paper, the authors attempt to compare and contrast two synchronic varieties of Sora, a South Munda language of the Austroasiatic language family, as it is spoken in Orissa, in eastern India, and Assam in northeastern India. This paper concludes that synchronically the Sora of Orissa and Assam have similar phonetic properties, the phonetic features of transplanted Sora, spoken in Assam, are preserved in a different linguistic region, even after hundreds of years of migration. This is a significant finding because it has been argued that transplanted varieties of a language become simplified by losing their idiosyncratic features after they are transplanted to a new sociolinguistic environment.

Shailendra Mohan presents the last article in this volume on "Noun Morphology in Korku". The article presents the number, gender, and case encodings in the Korku language. It also presents the noun derivation system, numerals, postpositions, and interrogatives in Korku. The description will help to fill the gap that exists about the knowledge of this language as well as to provide data for the comparative study of the other South Asian languages.

I sincerely hope teachers, students, and researchers of linguistics, especially South Asian Linguistics, will find this volume useful.

Editor
Shailendra Mohan

INTRODUCTION

MUNDA LANGUAGES: AN OVERVIEW[1]

SHAILENDRA MOHAN
AND MASATO KOBAYASHI

1.0 Introduction

Munda languages constitute the western branch of the Austro-Asiatic linguistic phylum. There are over 10 million speakers of Munda languages in India, living in an area stretching from the western part of the country, i.e. from Maharashtra to the Northeast of India. Munda speaking people live mainly in the states of Orissa and Jharkhand; significant Munda language speaking groups are also found in the states of Madhya Pradesh, Maharashtra, Andhra Pradesh, and Chhattisgarh and through migration to virtually all areas of India. Santhali is the only Munda language that has official status in India. It is included in the VIIIth Schedule of the Indian Constitution. Other Munda languages have no official status.

Munda languages have interacted with most of the other major language groups of India over several millennia and have logically both influenced and been influenced by various other families of languages of South Asia, e.g., Dravidian (Bhattacharya 1975a, Anderson 2003). Further, the Munda languages have their linguistic cousins to the east, so they also have features reflecting their shared history with various language groups of Southeast Asia from an earlier historical period. (Anderson 2014).

Figure i-1: Map of the Munda languages
(Anderson 2007:7, reprinted with kind permission of the author)

2.0 Munda Languages and its distribution

Grierson's Linguistic Survey of India is one of the earliest documents to provide information on Munda languages. On the whole, Grierson (1967) identifies 14 Munda languages. The whole Munda branch has been divided into a group of dialects. Grierson claims that "Kherwari is the principal Munda language, its dialects having been returned by full 88% of all the speakers of Munda tongues" (Grierson 1967:21). According to him, Kherwari is also the only Munda form of speech that has remained comparatively free from the influence of neighbouring languages.
'Kherwari' is a hypothetical language incorporating most of the northern Munda dialects. The names of 14 Munda dialects are as follows:

Sl.No.	Name of the Munda Dialect	Census of 1901
1	Santālī	1,795,113
2.	Muṇḍārī	400,744
3.	Bhumij	111,304
4.	Birhāṛ	526
5.	Koḍā	23,878
6.	Hō	371,860
7.	Tūrī	3,880
8.	Asurī	4,894
9.	Korwā	16,442
10.	Kūrkū	87,657
11.	Khaṛiā	82,506
12.	Juāng	10,853
13.	Savara	157,136
14.	Gadabā	37,230
	Total	3,164,036

Table i-1 Munda Languages in Grierson's Linguistic Survey of India

According to the 1961 Census report, 58 mother tongues have been found as belonging to the Munda Branch. The name of a language and its dialects are presented below:

Sl.No.	Mother tongues of Munda Languages	Grouping	Speakers
1.	Kherwari		647
2.	Santhali		3,130,829
3.	Grouped under Santhali	Gayari	16
4.		Gora	1
5.		Har	9
6.		Kamari-Santhali	903
7.		Karmali	90,849
8.		Kisan-Santhali	41
9.		Lohari-Santhali	130
10.		Mahili	19,697
11.		Manjhi	2,296
12.		Paharia	2,287
13.	Mundari		736,524
14.	Grouped under 'Mundari'	Mura	513

15.	Bhumij		1,31,258
16.	Grouped under ‘Bhumij’	Bhuiya/Bhuyan	10
17.		Kisan-Bhumij	
18.		Kurmi	351
19.		Larka	32
20.		Parsi-Bhumij	4,754
21.		Rahiya	1,029
22.	Birhor		590
23.	Koda/Kora		13,277
24.	Grouped under ‘Koda/Kora’	Khaira	18,325
25.		Mirdha-Koda/kora	76
26.		Udangmudia	46
27.	Ho		648,066
28.		Lohara	293
29.	Turi		1,562
30.	Asuri		4,540
31.	Agaria		98
32.	Birjia/Brijia/Binjhia		2,391
33.	Grouped under Birjia/Brijia/Binjhia	Pahrai-Birjia/Brijia/Binjhia	4
34	Korwa		16,286
35.	Grouped under ‘Korwa’	Jangali-Korwa	35
36.		Koraku	53
37.		Majhi-Korwa	1,339
38.		Singli	7
39.	Korku		208,165
40.	Grouped under ‘Korku’	Mankari	1,081
41.		Muwasi	9,829
42.		Nihali	1,167
43.	Kharia		171,269
44.	Grouped under ‘Kharia’	Baiti	5
45.		Dhelki	58
46.		Lodha	5
47.		Mirdha-Kharia	5,822
48.	Juang		15,795
49,	Savara		265,721
50.	Gadaba		40,193
51.	Munda-Unspecified		167,159
52.		Adibhasha Munda	13,140
53.		Kol	64,465

54.		Lohari-Munda	123
55.		Mahto	7
56.		Parenga	767
57.		Parhaiya	397
58.		Thar	15,595

Table i-2 Munda Languages according to 1961 Census of India

In the 1961 Census, 'Kherwari' was attested as a separate language because 647 speakers reported it as a mother tongue. In the 1971 Census report, 60 mother tongues have been found as belonging to the Munda Branch. Ten (10) mother tongues of the 1961 list failed to reappear in 1971 while 12 new mother tongues were reported in 1971 the census report (Mahapatra 1991:333). According to him, "the problem basically is to distribute these mother tongues into the language grid of the Munda group". After the 1971 Census, the artificial ceiling of less than 10,000 was applied, i.e. language speakers whose numbers are less than 10,000 were referred to in a separate tabulation. This criterion has wiped out many tribal languages, including the Munda languages whose population is less than 10,000. Those who returned less than 10,000 speakers each at an all India level were included in "Others". The Munda languages in the Census of 2011 are listed below:

Sl.No.	Name of the Munda Language	Languages included	No. of speakers
1.	Santhali		7,368,192
		Karmali	358,579
		Mahili	26,399
		Santali	6,973,345
		Others	9,869
2.	Bhumij		27,506
		Bhumij	10,190
		Others	17,316
3.	Gadaba		40,976
		Gadaba	40,965
		Others	11
4.	Ho		1,421,418

		Ho	1,410,996
		Lohara	10,422
5.	Juang		30,378
		Juang	30,378
6.	Kharia		297,614
		Kharia	293,665
		Others	3,949
7.	Koda/Kora		47,268
		Koda/Kora	47,181
		Others	87
8.	Korku		727,133
		Korku	688,053
		Muwasi	35,827
		Others	3,253
9.	Korwa		28,453
		Koraku	16,154
		Others	12,299
10.	Munda		505,922
		Kol	19,868
		Munda	464,817
		Others	21,237
11.	Mundari		1128,228
		Mundari	1128,050
		others	178
12.	Savara		409,549
		Savara	409,481
		Others	68

Table i-3 Munda Languages according to 2011 Census of India

The comparative growth of Munda Languages is listed below:

Comparative growth of Munda languages of India 1971, 1981, 1991 and 2001
(Census of India, Statement 8)

Sl. No.	Name of the Language	1971	1981	1991	2001	Decadal Percentage		
						1971-1981	1981-1991	1991-2001
1	Bhumij	51,651	50, 384	45,302	47,443	-2.45	-10.09	4.73
2	Gadaba	20,420	28,027	28,158	26,262	37.25	0.47	-6.73
3	Ho	751,389	783,301	949,216	1,042,724	4.25	21.18	9.85
4	Juang	12,172	19,038	16,858	23,708	56.41	-11.45	40.63
5	Kharia	191,421	212,605	225,556	239,608	11.07	6.09	6.23
6	Koda/ Kora	14,333	23,113	28,200	43,030	61.26	22.01	52.59
7	Korku	307,434	347,661	466,073	574,481	13.08	34.06	23.26
8	Korwa	15,097	48,079	27,485	34,586	218.47	-42.83	25.84
9	Munda	309,293	377,492	413,894	469,357	22.05	9.64	13.40
10	Mundari	771,253	742,739	861,378	1,061,352	-3.70	15.97	23.22
11	Santali	3,786,899	4,332,511	5,216,325	6,469,600	14.41	20.40	24.03
12	Savara	222,018	209,092	273,168	252,519	-5.82	30.64	-7.56

Table i-4. The Comparative growth of Munda Languages

Munda languages according to Ethnologue 2017[2]

	Language	Alternate Names	Population	Location
1.	Agariya	Agaria, Agharia, Agoria	72,000 (2007)	Chhattisgarh state: Bilaspur district; Madhya Pradesh state: Mandla and Rewa districts, Maikal hills; Uttar Pradesh state: Agra, Mathura, and Mirzapur districts.
2.	Bijori	Binjhia, Birijia, Birjia, Brijia, Burja	25,000 (1998 GRN)	Jharkhand state: Cowerdaga, and Ranchi districts; Madhya Pradesh and Odisha states; West Bengal state: Darjeeling and Jalpaiguri districts.
3.	Kodaku	Koraku, Korwa	15,700 (1991 census)	Chhattisgarh state: Surguja district; Jharkhand state: Garhwa and Palamau districts; Uttar Pradesh state: Sonbhadra district.
4.	Asuri	Ashree, Assur, Asura, Maleta	7,000 (Van Driem 2007)	Chhattisgarh state: Raigarh district, Jashpur area; Jharkhand state: Gumla, Lohardaga, southern Palamau, and northern Ranchi districts of Chotanagpur Plateau; Maharashtra state; Odisha state: Sambalpur district; West Bengal state.
5.	Birhor	Birhor, Birhar, Birhore, Birhul, Mankidi, Mankidia, Mankiria	2,000 (Van Driem 2007)	Chhattisgarh state: Raigarh district; Jharkhand state: southern Hazaribag, southern Palamau, Ranchi, and Singhbhum districts; Maharashtra state; Odisha state: Kalahandi, Keonjhar, Mayurbhanj, Sambalpur, and Sundargarh districts; West Bengal state: Puruliya district.

6.	Ho	Bihar Ho, Lanka Kol	1,040,000 (2001 census)	Jharkhand state: PurbiSinghbhum district, Kolhan, Seraikella; East Singhbhum district, Dhalbhum sub-district; Odisha state: Koenjhar and Mayurbhanj districts; West Bengal state.
7.	Koda	Kaora, Kora, Korali, Korati, Kore, Mudi, Mudikora	43,000 (2001 census)	West Bengal state: Bankura and Bardhaman districts.
8.	Kol	Hor	1,660 (2012 SIL)	Rajshahi district: Godagari subdistrict.
9.	Korwa	Ernga, Singli	34,600 (2001 census)	Chhattisgarh state: Bilaspur, Jashpur, Korba, Raigarh, and Surguja districts; Jharkhand state: Gumla, Garhwa, and Palamau districts; Odisha state: Mayurbhanj and Sundargarh districts; Uttar Pradesh state: Mirzapur district; Andhra Pradesh, Maharashtra, and West Bengal states
10.	Munda	Heriki, Killi	469,000 (2001 census)	Odisha and Jharkhand states; possibly Bihar and West Bengal.
11.	Mundari	Colh, Horo, Kolh, Mandari, Mondari, Munari	1,110,000 (2001 census)	Jharkhand state: Ranchi district, south, and west; Andaman and Nicobar Islands union territory, Assam, Himachal Pradesh, Madhya Pradesh, Odisha, Tripura, and West Bengal states.

12.	Mahali	Mahili, Mahle, Mahli	33,000 (2007)	Assam state: tea estates; Jharkhand state: Dhanbad, Gumla, Hazaribagh, Pargana, Ranchi, SantalLohardaga, SaraikelaKharsawan, East Singhbhum, and West Singhbhum districts in Chota Nagpur area; Odisha state: Balasore, Keonjhar, and Mayurbhanj districts; West Bengal state: Jalpaiguri and West Medinipur districts.
13.	Santhali	Har, Hor, Samtali, Sandal, Sangtal, Santal, Santali, Santhiali, Satar, Sentali, Sonthal	5,940,000 (2001 census)	Bihar state: Bhagalpur and Munger districts; Jharkhand state: Hazaribagh and Manbhum districts; Odisha state: Balasore district; West Bengal state: Bankura and Birbhum districts; Assam, Mizoram, and Tripura states.
14.	Turi		2,000 (2007)	Chhattisgarh state: Raigarh district, and scattered throughout; Jharkhand state: Gumla, Lohardaga, and Ranchi districts, Chotanagpur area; Odisha state: Sambalpur and Sundargarh districts; West Bengal state: Bankura, Birbhum, Murshidabad, and Nadia districts.
15.	Korku	Bondeya, Bopchi, Korki, Kuri, Kurku, Kurku-Ruma, Ramekhera	574,000 (2001 census)	Madhya Pradesh state: Betul district, Betul city area and north; Hoshangabad and East Nimar (Khandwa) districts; Maharashtra state: Akola, Amravati, and Buldana districts.

16.	Juang	Juango, Patra-Saara, Patua, Puttooas	23,700 (2001 census)	Odisha state: north Angul, east Dhenkanal, south Keonjhar districts.
17.	Kharia	Haria, Khadia, Khariya, Kharvi, Khatria, Kheria	240,000 (2001 census)	Jharkhand state: Ranchi district, Khunti sub-district, Kolebira and Thethaitangar Anchal; Simdega sub-district; Bilaspur, Chhattisgarh, Durg, Jashpur, Raigarh, Raipur, East Singhbhum, and West Singhbhum districts; Odisha state: Mayurbhanj, Sambalpur, and Sundargarh districts; Andaman and Nicobar Islands, Assam, Tripura, West Bengal states. Dhelki dialect mainly in northwest Gangpur (Raigarh), Jashpur, and Sundargarh; Dudh dialect in south Gangpur (Raigarh) Ranchi, and western Sambalpur.
18.	Gata’	Didayi, Didei, Dire, Gataq, Geta’, Getaq, GtaAsa, Gta’	3 06 (1991 census)	Andhra Pradesh state: east Godavari district; Odisha state: Koraput and Malkangiri districts, Kudumulgumma and Chitrakonda sub-districts south of Bondo Hills; some in Khairput sub-district. 47 villages.
19.	Bondo	Bhonda Bhasha, Bonda, Bondo-Poraja, NanqaPoroja, Poraja Katha, Remo, Remosum	9,000 (2002 SIL)	Odisha state: Malkangiri district, Khoirput sub-district, Bondo Hills.

20.	Gadaba, Bodo	BoiGadaba, Gadba, Gadwa, Godwa, Gudwa, Gutob, Gutop	8,000 (2000 IICCC)	Andhra Pradesh state: Visakhapatnam district; Odisha state: Koraput district, Lamtaput sub-district, 40 villages; Malkangiri district, Khoirput sub-district.
21.	Parenga	Gadaba, Gorum, Gorum Sama, Pareng, Parenga Parja, Parengi, Parenji, Poroja	12,600 (2001 census)	Odisha state: Mayurbhanj district.
22.	Juray		801,000 (2000)	Odisha state.
23.	Sora	Sabar, Sabara, Saonras, Saora, Saura, Savara, Sawaria, Shabari, Soura, Swara	253,000 (2001 census)	Andhra Pradesh state: Srikakulam district; Assam state: Plains division; Odisha state: Ganjam, Koraput, and Phulbani districts; Bihar, Madhya Pradesh, Tamil Nadu, and West Bengal states.

Table i-5. Munda languages according to Ethnologue 2017

3.0. Munda and Austroasiatic

Munda languages represent the westernmost branch of the Austroasiatic linguistic phylum. Austroasiatic (AA) languages are geographically widespread from Central India to Southeast Asia occurring in discontinuous pockets of the speech communities. Vietnamese and Khmer are the national languages of Vietnam and Cambodia, respectively. Santali is the language listed in the Schedule VIII of the Indian Constitution, and Khasi has official status in the state of Meghalaya, India. The other Austroasiatic languages have no official status. It was Wilhelm Schmidt in 1906 who established the existence of the Austroasiatic group. According to Jenny, Weber, and Weymuth (2015:13):

> "Typologically, the AA languages can be superficially grouped into three distinct subgroups which belong to three geographic regions of the AA speaking area. The Munda languages in Central and Eastern India are consistently verb-final agglutinating languages, with a large number of affixes expressing derivational processes as well as case relations with nominals, and tense-aspect and person with verbs. The Nicobarese languages, spoken on the Nicobar Islands in the Andaman Sea, are generally verb-initial and exhibit complex morphological processes, including prefixes, infixes, and suffixes. The rest of the family, mostly spoken in Mainland Southeast Asia, is generally verb-medial, and apart from traces of inherited derivational morphology, isolating".

This diversity in the structural type of the Austroasiatic language family led Grierson (1906:2) to remark that if they were descended from a common language, the language must have been adopted by people with opposite orders of thought. The genetic relationship of the Indian and Southeast Asian members of the Austroasiatic language has been established (Schmidt 1906, Pinnow 1959, etc.). The classification of Austroasiatic languages by Pinnow (1959) is as follows:

A. West-Nordwest: Nahali

B. Nordwest: Munda

a) Ost: Kherwari
b) West: Kurku
c) Zentral: Khaɽia- ɟuaŋ
d) Süd: Sora-Gadaba

C. Mon-Khmer

a) Mon
b) Kambodja (e.g. Khmer, Pear, Stieng)
c) Chəma
d) Mnong
e) Bahnar
f) Sedang
g) Brao
h) Jəru
i) Kuoy
j) Suoy (e.g. Kaseng, Alak, Laveh)

D. Nordost: Palaung-Wa, or Salwen Group
a) West (e.g. Riang, Palaung, Wa etc.)
b) Ost (e.g. Khmu, Lamet etc.)

E. Nord: Khasi
F. Südwest: Nikobar
G. Səmang
H. Sakai
I. ɟaku'd (- ɟakud'n, - ɟakun)

Figure i-2 Austroasiatic languages by Pinnow (1959)

The most recent assessment of Austroasiatic classification is that of Sidwell (2009), who situates all Austroasiatic languages with an equal constituent.

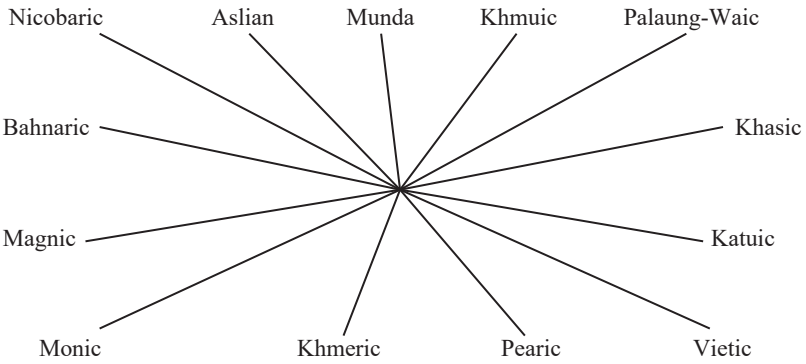

Figure i-3: Sidwell's (2009) model of Austroasiatic (cited in Anderson, 2015)

Researchers who are working on comparative Austroasiatic have been troubled by the apparent typological opposite structures in Mon-Khmer languages on the one hand, and the Munda languages on the other. Some consider the Munda languages to have acquired structures as a result of diffusion from the neighbouring Dravidian and Indo-Aryan languages, including the verb-final constituent order (Donegan, 1993; Donegan & Stampe 1983). Others, e.g. Pinnow (1963), Anderson & Zide, 2001) suggest that Proto-Austroasiatic was more like Munda, with rather extensive morphological complexity. Donegan & Stampe (2004:1) present

the typological opposition between Munda and Mon-Khmer. They argue that their opposite synthetic vs analytic traits might be explained as due to polar drifts driven by their opposite falling vs rising phrase and word rhythms.

	Munda	Mon-Khmer
Phrase Accent	Falling (initial)	Rising (Final)
Word Order	Variable-OV,AN, postpositional	Rigid-VO,NA, prepositional
Syntax	Synthetic-subj/obj agreement on the verb	Analytic-no inflectional morphology
Word Canon	Trochaic	Iambic, monosyllabic
Morphology	Agglutinative, suffixing, polysynthetic	Fusional, prefixing or isolating
Timing	Isomoraic or isosyllabic	Isoaccentual
Syllable Canon	(C)V(C)	Unaccented (C) ə, accented (C) (C)V (G)(C)
Consonantism	Stable, geminate clusters	Shifting, tonogenetic, non-geminate clusters
Tone/ Register	Level tone (Korku only)	Contour tones or registers
Vocalism	Stable, monophthongal, harmonic	Shifting, diphthongal, reductive.

Table i-6 Munda and Austroasiatic (Donegan & Stampe 2004)

4.0. Sub-groupings of Munda Languages

There is a consensus that Munda languages can be divided into mainly two groups, namely North Munda and South Munda (for an alternate view see Anderson, this volume). Pinnow (1959:1-3) presents the first classification of Munda languages in his landmark study on historical phonology of the Kharia language.

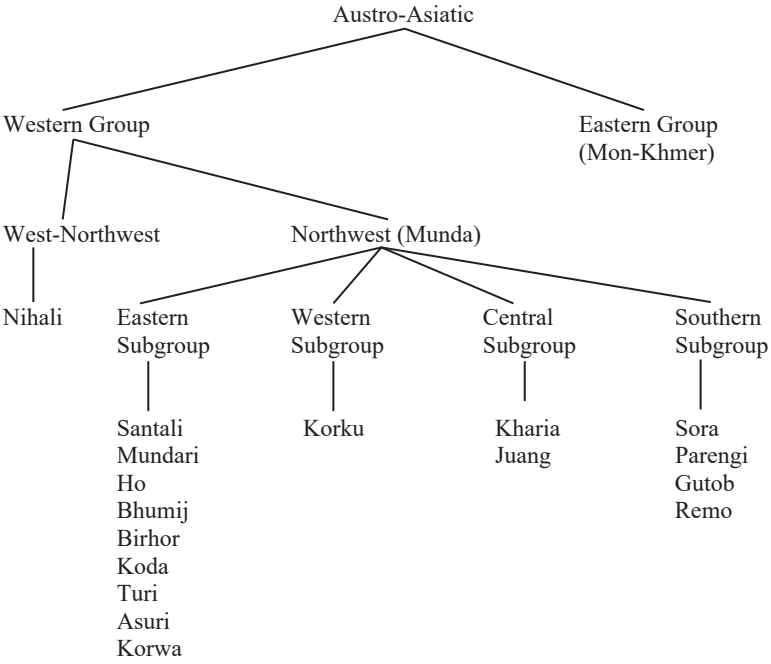

Figure i-4 The Munda languages and Nihali, according to Pinnow (1959:1-2)

After ten years, Zide (1969:412) presented a different classification of Munda languages. Pinnow's (1959) classification of "Eastern Subgroup" and "Western Subgroup" of Munda has been combined under the term "North Munda". Pinnow's "Central Subgroup" and the "Southern Subgroup" now together form the "South Munda". Zide identifies ten languages: Sora, Gorum, Gutob, Remo, Gata?, Kharia, Juang, Korku, Santhali, and Mundari-Ho. Zide's classification has become the most widely accepted and is now generally viewed as the "Traditional classification of the Munda Languages" (Anderson,2008:2).

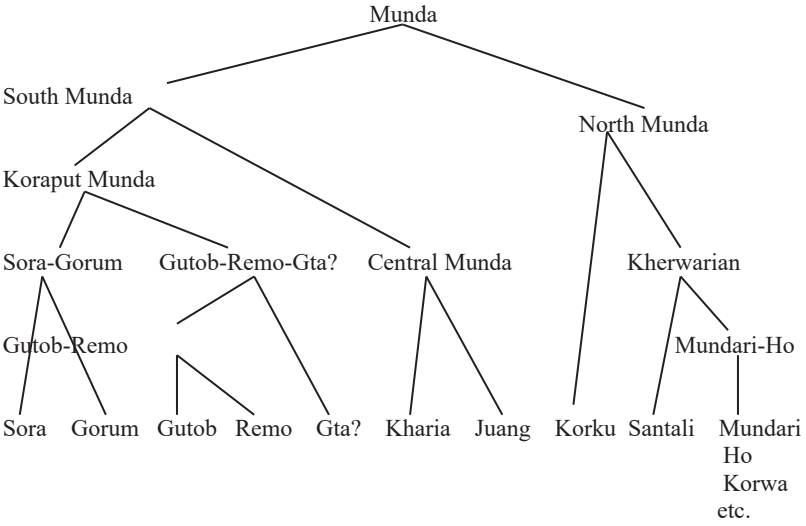

Figure i-5: The Munda languages, according to Zide (1969:412).

Sudhibhusan Bhattacharya proposed the other configuration in 1975. The Munda languages are ten in number and usually classified into the following sub-branches:

(1) Northwestern Munda, consisting of Korku and its dialect Mowasi;
(2) Northern Munda, consisting of Kherwari, i.e. Santali, Mundari, and their dialects;
(3) Central Munda, consisting of Kharia (Kheṛia) and Juang;
(4) Southern Munda, consisting of Saoṛa (= Sōra), Parengi (= Gorum), Gutob, Bonḍa (= Remo), and Ḍiḍey (= Gtaʔ)

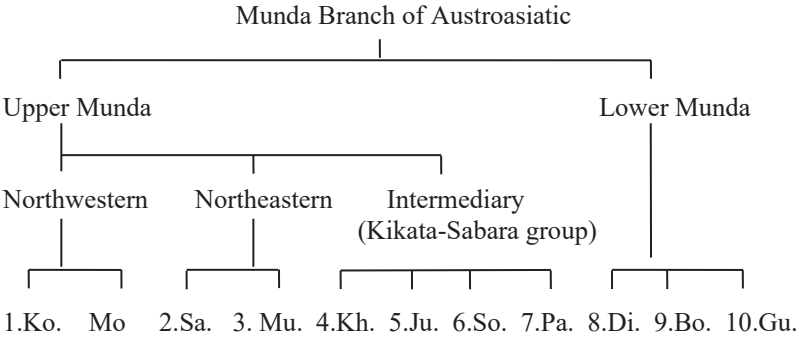

Figure i-6: Munda language relationship by Bhattacharya (1975)

The other important revision to Munda classification was proposed by Anderson (1999) in which South Munda is directly divided into three daughter groups, namely Sora-Gorum, Kharia-Juang, and Gutob-Remo-Gtaʔ, i.e., without the "Koraput Munda" group assumed in Zide (1969).

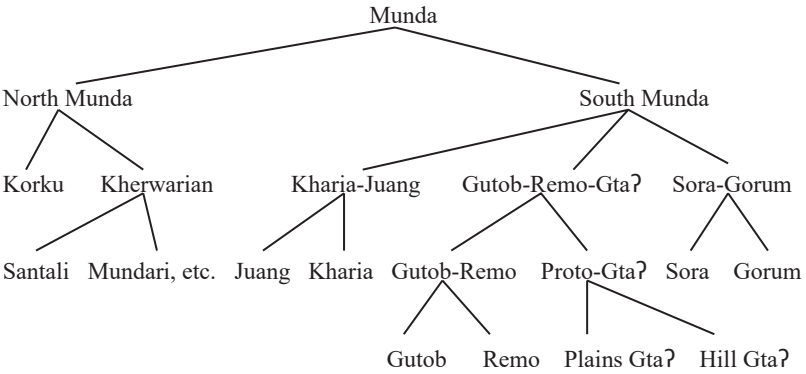

Figure i-7: The Munda languages, according to Anderson (2001), cited in Anderson (2007, 2008a:4).

A new Classification of Munda is presented in this volume by Gregory D.S. Anderson (For details see Chapter One).

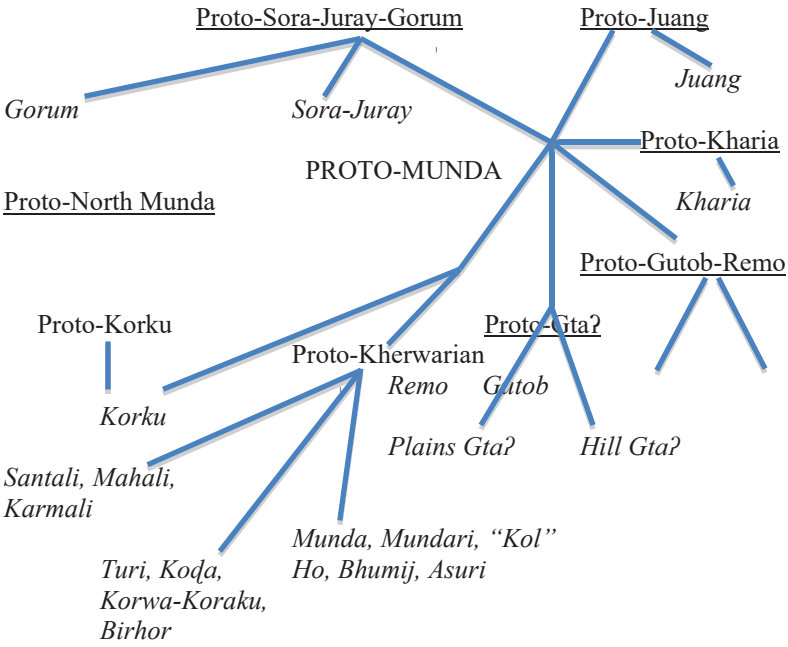

Figure i-8 Classification of Munda Languages by Anderson (this volume)

5.0 Summing up

This chapter provides the issues related to Munda languages, their distribution in the Indian subcontinent, major groupings, and their relationship with Austroasiatic languages. It is hoped that this overview will help future researchers to understand the issues in Munda linguistics.

Notes

[1] I am thankful to the ICSSR (India)-JSPS (Japan) bilateral joint research Project on "Establishment of International Munda Network" for the support.
[2] I am thankful to Luke Horo for preparing this chart for me.

References

Anderson, Gregory D. S. "A New Classification of Munda: Evidence from Comparative Verb Morphology." *Indian Linguistics* 62, (2001): 27-42.

—. "Dravidian influence on Munda." *International Journal of Dravidian linguistics* 32, no.1 (2003): 27–48

—. "Advances in Proto-Munda Reconstruction." *Mon-Khmer Studies* 34 (2004): 159-184.

—. *The Munda Verb. Typological Perspectives.* Berlin: Mouton de Gruyter (Trends in Linguistics, Studies, and Monographs, 174), 2007.

—, ed. *The Munda Languages*. London & New York: Routledge, 2008.

—. "What Munda languages are really like." Keynote address presented at the All India Conference of Linguists (AICL-34), Shillong, India, November 1-3, 2012. Keynote address.

—a. "Overview of the Munda languages." In *The Handbook of Austroasiatic Languages*, *Volume 1*, edited by Mathias Jenny and Paul Sidwell, 364-414. Amsterdam: Brill, 2014.

—b. "Prosody, phonological domains and the structure of roots, stems and words in the Munda languages in a comparative/historical light". *Journal of South Asian Languages and Linguistics*, no.2 (2015): 163-183.

Anderson, Gregory D. S., and Bikram Jora. "Negation, TAM and person-indexing interdependencies in the Munda languages: a preliminary report". In *Papers from the Seventh International Conference on Austroasiatic Linguistics*, edited by Hiram Ring and Felix Rau, 36-59, Hawai: University of Hawai'i, 2018.

Bhattacharya, Sudhibhushan. "A new classification of Munda". *Indo-Iranian Journal* 17 (1975): 97-101.

Census of India. Table C-16, Office of the Registrar General, India, 1961

Census of India. Table C-16, Office of the Registrar General, India, 2011

Donegan, Patricia J. "Rhythm and Vocalic Drift in Munda and Mon-Khmer". *Linguistics in the Tibeto-Burman Area* 16, no.1 (1993): 1-43.

Donegan, Patricia J., and David Stampe. "Rhythm and the Holistic organization of Language Structure". In *Papers from the Parasession on the Interplay of Phonology, Morphology, and Syntax*, edited by John F. Richardson, Mitchell Marks, and Amy Chukerman,337-353 Chicago: Chicago Linguistic Society,1983.

—. "Rhythm and synthetic drift of Munda". In *Yearbook of South Asian Languages and Linguistics* edited by Rajendra Singh, 3-36, Berlin, Mouton, 2004.

Grierson, G.A. *Linguistic Survey of India*, Vol. 4, Munda and Dravidian. New Delhi: Low Price Publications, 1967[First published1906].

Jenny, M, Tobias Weber, and Rachel Weymuth. "The Austroasiatic Languages: A Typological Overview". In *The Handbook of Austroasiatic Languages*, *Volume I*, edited by Mathias Jenny and Paul Sidwell, 13-143. Amsterdam: Brill, 2015

Mahapatra, B.P. "Munda Languages in Census". *Bulletin of the Deccan College Post-Graduate and Research Institute*, 51-52 (1991):329-336.

Pinnow, Heinz-Jürgen. 1959. *Versuch einer historischen Lautlehre der Kharia-Sprache* Wiesbaden: Otto Harrassowitz.

Schmidt, Peter P. W. "Die Mon-Khmer Völker: Ein Bindeglied zwischen Völkern Zentralsiens und Austronesiens". *Archiv der Anthropologie, Braunschweiger* 33 [n.s. 5]: 59–109, 1906.

Simons, Gary, F, and Charles D Fenning. *Ethnologue: Languages of Asia*. Summer Institute of Linguistics, 2017.

Zide, Norman H. "Munda and non-Munda Austroasiatic languages". In *Current Trends in Linguistics 5,* edited by Thomas Sebeok, 411–430. The Hague: Mouton, 1969.

CHAPTER ONE

ON THE ROLE OF AREAL AND GENETIC FACTORS IN THE DEVELOPMENT OF THE WORD-STRUCTURE AND MORPHOSYNTAX OF THE MUNDA LANGUAGES

GREGORY D. S. ANDERSON

1.0 Introduction

I offer here some comments on the varied roles of areal and genetic (or 'external' vs 'internal') factors in the development of the word structure and morphosyntax of the Munda languages. I begin with a brief assessment of the current state of Munda language classification in section 2. The remainder of the paper presents arguments for a more nuanced approach to the role of contact vs. inheritance in the historical analysis of the Munda languages. Specifically, some features are atypical of South Asia found in Munda, some of which are ancient features of Munda from the Austroasiatic. These are discussed in section 3. In other instances, there are clearly secondarily acquired features of Munda languages that *do* speak to their extensive contact history with the dominant South Asian language families, a selection of these are presented in section 4. However, contact-triggered changes also appear to have applied to different domains in different Munda languages at distinct historical periods, and thus a one-time parametric shift proposed by the theory of rhythmic holism (Donegan 1993, Donegan and Stampe 1983, 2004) that served to change Munda languages from canonically Southeast Asian to canonically South Asian cannot be maintained. There are also features of Munda which might reflect pre-Aryan-Dravidian contact strata in the history of Munda, or otherwise lack analogs in other Austroasiatic groups and in Indo-Aryan and Dravidian languages, but are reminiscent of features found in the periphery of South Asia now, in Munda, Tibeto-Burman (Trans-Himalayan)

and in the isolate language Burushaski, which might reflect an earlier pre-Arayo-Dravidian profile of South Asia attested in Munda as well. This is presented in section 5. Finally, in section 6, I discuss an area of Munda grammar where we see archaic retentions, contact-driven restructuring, and internal developments all separately manifested but interacting in the patterns of inflection seen across the Munda languages in auxiliary verb constructions.

2.0 The Internal and External Classifications of Munda

Until the second decade of the 21st century, a lot of potential progress in the historical analysis of Austroasiatic was hamstrung by an enduring misconception that there was a primary split between Munda and all other groups, traditionally labelled 'Mon-Khmer' (Shorto 1976, Pinnow 1959, 1960, 1963, Diffloth 1991, Shorto 2006, Sidwell 2008). Recent work in the phylogenetic analysis of Austroasiatic (Sidwell 2009, 2011, 2013, 2015a) has suggested that Austroasiatic does not have a hierarchical structure but rather is a wheel with thirteen equal branches, one being Munda (Figure 1).

The standard classifications of Munda tend to also recognise primary splits, either a North [Korku-Kherwarian] vs. South [the rest] one (Zide 1969), or Lower [Gutob-Remo-Gtaʔ] vs Upper [the rest] one (Bhattacharya 1975); earlier classifications (Pinnow 1959) recognised four coordinate branches: Korku, Kherwarian, Kharia-Juang, Koraput Munda. Recent work (Anderson 2015, 2016, Sidwell and Rau 2015) has questioned all these approaches, and recognizes six coordinate branches, as in Figure 2: Kharia, Juang, and Gtaʔ form isolate branches, Zide's North Munda is retained as such (Korku-Kherwarian), and Sora-Gorum and Gutob-Remo form small branches coordinate with all the previously named branches.

Although I do not attempt reconstructions here, methodologically speaking, I consider this to mean that any (non-copied) feature/form found in any three (physically) non-adjacent groups should *a priori* be entertained as a possible proto-language feature/form. This holds for Munda-internal and pan-Austroasiatic comparisons alike. If a feature/form is found in only two non-adjacent groups, it should still be considered a possible candidate as attestations of a potential inheritance from a proto-language feature/form.

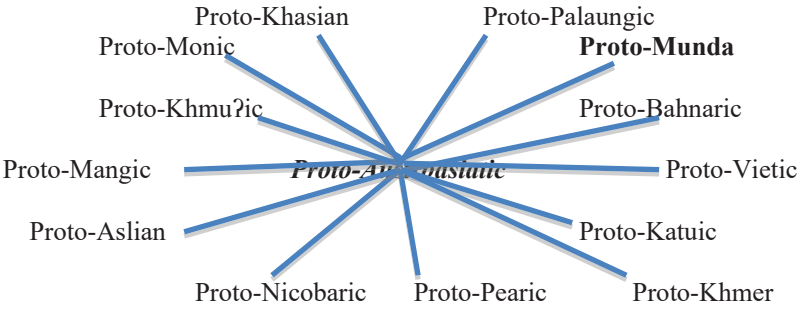

Figure 1-1: Classification of Austroasiatic (Sidwell 2015a)

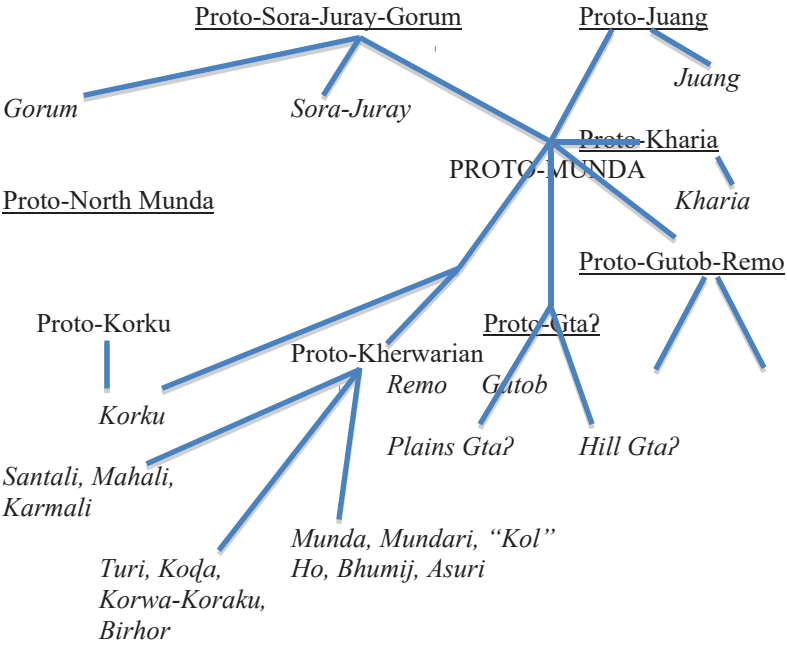

Figure 1-2: Classification of the Munda languages using lexical and grammatical data (Anderson 2016ms)

3.0 Features of Munda not reflecting contact with major South Asian Languages

A number of features found in present-day Munda languages clearly reflect secondary accommodation to South Asian areal norms. These are discussed in 4 below. However, the role of contact with Indo-Aryan and Dravidian languages, while indeed pronounced in individual languages and in specific domains across the Munda languages (use of retroflex consonants-Arsenault 2012, 2017), has been overstated in the literature. Indeed, what has been claimed is, at times, incorrect and misleading (Ring and Anderson 2018). Foremost among these is the alleged operation of a principle called 'rhythmic holism' (Donegan 1993, Donegan and Stampe 1983, 2004), which states that a one-time reset of the rhythm of Munda languages from rising (iambic) to falling (trochaic) triggered a series of different changes across different domains of the language–presumably Proto-Munda for the supposed changes to have uniformly applied across all domains in all Munda languages, ranging from phonetics to syntax. Some issues with why rhythmic holism should be abandoned are addressed in 3.1 and 3.4. In the remaining subsections here we discuss three features of Munda languages that retain older features found in other Austroasiatic languages. In 3.2, we examine a system of prefixes used with nouns across the family that may speak to an earlier system of noun classification in the Austroasiatic group. In 3.3, examples of pre-verbal negation in Munda are offered with parallels in eastern Austroasiatic languages. In 3.4, the syntax and use of classifiers in Munda are discussed with reference to cognate systems found in other Austroasiatic languages, and we see a cline of accrual of South Asian areal features in the NP syntax of the Austroasiatic groups of South Asia.

3.1 On the fallacy of Rhythmic Holism: Prosodic Features and Munda Morphosyntax

A belief in a false dichotomy between Munda and 'Mon-Khmer' has had considerable consequences for the advancement of the study of the history of the Munda languages. This is perhaps best embodied by the following quote that Munda and 'Mon-Khmer' are "systematically *opposite* at every level" (Donegan and Stampe 2004 [DS04]: 3, 5). Among the criteria used to support this claim are included those in (1).

(1)

Domain	Munda/Indosphere		Mon-Khmer/Sinosphere
Grammar	"Synthetic"	vs.	"Analytic"
Words	Falling (Trochaic)	vs.	Rising (Iambic)
Consonants	Stable/Assimilative	vs.	Shifting/Dissimilative
Vowels	Harmonizing/Stable	vs.	Reducing/Diphthongizing

Belief in a primary taxonomic split between Munda and all other related languages coupled with a belief that there was a contact-triggered, complete restructuring of Munda at the proto-Munda level entails from the perspective of the historical methodology of comparative Austroasiatic linguistics until quite recently, that everything in Munda is diachronically secondary, that it straightforwardly reflects accommodation on every level from phonology to syntax to morphology, etc., to South Asian areal norms, and thus Munda data are of little historical importance to the study of the Austroasiatic phylum as a whole as they largely shed no light whatsoever on the history of other Austroasiatic groups. All of these misconceptions are due to the supposed operation of a unifying principle of rhythm that was reset from an older Austroasiatic setting (Mon-Khmer/Sinosphere) to a newer South Asian one (Munda/Indosphere), a theory called rhythmic holism. The hypothesis of rhythmic holism has only recently been seriously scrutinized (Horo and Sarmah 2015, Horo 2017, Anderson 2015, 2020, Ring and Anderson 2018), although similar sentiments were echoed in Sidwell (2012) and Jenny et al. (2015). The hypothesis asserts that due to an alleged shift in the 'rhythmic holism' from iambic to trochaic at the proto-Munda level, a full shift to so-called 'Indospheric' and 'synthetic', head-final norms followed, away from the original 'Sinospheric'/'analytic', head-initial ones. However, no such immediate parameter re-setting occurred that simultaneously triggered a restructuring of the prosody, morphology, and syntax of proto-Munda that then was inherited by all the Munda languages. Rather, some restructuring seems to have occurred at the proto-language level, at the level (of some features) of intra-clausal and sentential syntax, and some phonological characteristics (full phonological vowels in initial syllables becoming dominant), but that this shift to Indospheric norms in the morphology, phonology, prosodic domains, and indeed other areas of the syntax has occurred at different times and differently in individual Munda languages, suggesting this accrual of South Asian features is ongoing, and has yet to reach the complete restructuring into the mirror-image type or inverse of Sinospheric norms at every level that the proposal of rhythmic holism

entails, much less even at the prosodic level (Ring and Anderson 2018), and as such this hypothesis must be abandoned. Indeed, Munda languages remain morphotactically and phonologically more similar to other Austroasiatic languages than previously realised (Anderson 2020, Ring and Anderson 2018); moreover, when undoing Sinospheric metalinguistic analysis filters, greater morphological complexity in earlier stages of Austroasiatic languages of MSEA will be revealed than previously appreciated.

For example, many MSEA Austroasiatic languages do not lack any traces of synthesis, but rather synthetic structures can be found in wave-like clines of variability in essentially all branches of the family (Alves 2013, 2014, 2015). Synthetic structures of some sort or another have thus been proposed for Nicobarese (Radhakrishnan 1981), Aslian (Omar 1975, Benjamin 1976, 2011, Burenhult 2002, Matisoff 2003a, Kruspe 2004), Khasian (Nagaraja 1993), Palaungic (Milne 1921), Khmuic (Svantesson 1983), Mangic (Li 1996, Li and Lou 2015), Khmeric (Bauer 1986, You Sey 1976, Schiller 1994, Thomas 1990), Monic (Jenny 2003, 2005, Bauer 1982, 1989), Katuic (Watson 1964, 1966; R. Watson 2011, Solntseva 1996, Costello 1966, 1998, 2001, Bauer 1990, Alves 2004), Bahnaric (Smith 1969, 1973, Gradin 1976, D. Thomas 1969, Bauer 1987-1988) and even Vietic (Enfield and Diffloth 2009, Alves 2003, 2005) if not Vietnamese itself of course.

Among the most synthetic of non-Munda, Austroasiatic languages must clearly be reckoned Nicobarese and Aslian, but many subgroups manifest synthesis in the word to some minor degree. Words with three morphemes can be found in Aslian Jahai.

(2) Jahai

t-b-tadɔʔ
REL-PROG-wait
'waiting'
(Kruspe, Burenhult & Wnuk 2015: 423)

Aslian Maniq has a number of quasi-inflected synthetic forms, typically instantiating aspectual or Aktionsart categories; morphotactically one finds elements varying between prefixes and infixes.

(3) i. Maniq ii. Maniq

ma/s/bas *bi.mi/t/kāt* < *bikāt*
PROG-IPFV-run cry/PROG/IPFV/cry
'to be running' 'to be crying'
(Kruspe, Burenhult & Wnuk 2015: 432)

iii. Maniq iv. Maniq v. Maniq vi. Maniq

l-wa *pi/l/ŋok* *l-k-cɔk* *ca/l//p/kip*
REP-go sit.so/REP/sit.so REP-IPFV-pierce bend.head.down PROG/IPFV/
'walk repeatedly' 'sit s.o. repeatedly''pierce repeatedly' 'bend head up/down repeatedly'
(Kruspe, Burenhult & Wnuk 2015: 432) <*cakip*

Car Nicobarese shows a number of morphologically complex structures. This can yield words with more than three morphemes as well.

(4) i. Car Nicobarese ii. Car Nicobarese iii. Car Nicobarese

han-t/an/fɤ̌t-ə *lɛ́h-hə-lə* *ha-rɔ́:n-haka*
action.PFX-sling/NMLZR/sling-SFX hit-INCORP.OBJ-UP CAUS-slant-CONT
'by slinging' 'hit (it) up' 'is standing slantwise.'
(Sidwell 2015b: 1239, 1261, 1262)

iv. Car Nicobarese

kɨhí:t-hɛ: *káha* *ʔəc* *vo:k* *ŋac* *havɤ́t-kə-rɛ*
finish-SEQ body 1SG.SUBJ bathe finish walk-IPFV-REFL
'immediately after bathing, I walked around.'
(Sidwell 2015b: 1262)

Palaungic Danaw shows inflectional forms as well, but here they belong mainly to the modal and polarity domains, as well as non-finite marking.

(5) i. Danaw ii. Danaw iii. Danaw iv. Danaw v. Danaw
pə-jɔx *ɛ̄tə-bɐ* *lɐ-kɨ̂-nɐʔ* *pɐ-p^{h}ɛ̄* *mə-puūt=nə*
NFIN-do OPT COND-disappear NEG-go-PHB ABIL-read IRR=come=TOP
'doing... 'may (it) disappear' 'don't go!' 'can read' '(before) it comes'
(Si 2015: 1110, 1116, 1118, 1121)

Khasian languages also reflect a mild degree of synthesis. So subject markers are cliticized to the future (and negative) in Khasi, and more than one voice prefix may be found on a single stem as well.

(6) i. Khasi ii. Khasi

pʰaʔ-pɨn-jap	*ki=n*	*wan*	vs.	*ki*	*la*	*wan*
CAUS-CAUS-die	3PL=FUT	come		3PL	PST	come
'make kill'	'they will come'			'they came'		

(Nagaraja 2015a: 1173)

In Pnar, a cliticized non-finite marker has developed that also can create polymorphemic structures when added to, for example, a causative marked verb.

(7) i. Pnar ii. Pnar

u=pn-jap	*ja-ɛm-bnta*
NFIN=CAUS-die	BEN-have-purpose
'to murder'	'it has purpose'

(Ring 2015: 1212-1213)

All of the above-mentioned languages are spoken either on the periphery or outside the core Mainland Southeast Asian area where most Austroasiatic languages are spoken. Languages more in the core area show less synthesis unsurprisingly, but even there it can be found in individual languages. Thus Pacoh (Alves 2015) has a case prefix (see below), while Sedang can add its adversative prefix even to disyllabic stems.

(8) Sedang

lə-kədɛ
ADVERS-kill
'(afraid) will kill'
(Smith and Sidwell 2015: 800)

Old Mon made use of an irrealis prefix, possibly reinforced by Thai contact (Jenny 2015) that yields a (semi-)inflected form like the following:

(9) Old Mon

si-kʔim
IRR-smile
'will/that/should smile'
(Jenny 2015: 536)

What this means is that most likely a small number of clitics or prefixes served as functional elements in Proto-Austroasiatic (perhaps no more than one per word was permitted) and that the radically isolating profile of modern-day Mainland Southeast Asia was not exactly the likely proto-type of the proto-language. Thus the mild degrees of synthesis seen on the peripheral areas of Austroasiatic are probably closer to the original profile. Munda clearly innovated here, but Proto-Munda does not seem to have been so radically distinct from its sister languages at the same period of time.

3.2 Proto-Munda syntactically free vs bound combining forms and prosodic vs morphological prefixes in minor or initial syllabism in Austroasiatic

Two very different types of data within Munda languages give insight into the possible earlier structure of Austroasiatic, but both involve the same element. One domain is the derivation of syntactically free-standing nominals from underlying roots that fail to meet minimal word constraints, and the other is from noun incorporation, which in turn shows these same underlying roots, the so-called combining forms, which by definition are bound to another root together with which will satisfy any minimal word constraint.

With regards to the syntactically freestanding forms of nouns in Munda languages, all groups but the Kherwarian languages (and even some of these do too) appear to require augmentation to the underlying root. Many of these involve old word-building processes, such as *–n-* infixation, reduplication, etc. In a small number of cases, such as in the words for 'broom' and 'turmeric (also yellow)', a cognate derivational process yielding the syntactically free-standing form is seen across all the Munda languages, but more typically there is a range of different strategies selected by individual languages with respect to individual specific roots. So 'bear' has prefixation in Sora-Gorum, Gutob-Remo, and Gtaʔ, it has *–n-*infixation in Kherwarian and *–n-* infixation plus suffixation/compounding

in Kharia and Juang. The word for 'hand' has glottal stop infixation in Sora-Gorum and possibly historically in Kharia (or glottal stop suffixation), but a syllabic nasal prefix (or its reflex) in Juang and one form in Gtaʔ and reduplication in Gutob-Remo and the other form in Gtaʔ (where it is possibly a Remo loan). Such discrepancies across the languages in the means of creating syntactically free forms of nouns utilizing clearly cognate combining forms or roots are the norm, not the exception.

Kherwari	Sora	Gorum	Kharia	Juang	Gutob	Remo	Gtaʔ	gloss
dʒɔnɔ[k']	*dʒənoʔ*	*dʒɔnɔ*	*dʒonoʔ*	*dʒenɔk*	*sunog*	*sunug*	*ʧɲoʔ*	'broom'

sasaŋ	*saŋsaŋ*	*saŋsaŋ*	*saŋsaŋ*	*sa(ʔa)ŋsaŋ*	*sasaŋ*	*saŋsaŋ*	*ssia*	'turmeric'
bana	*kəmbud*	*kibud*	*bane/ai*	*banae*	*gubɔn*	*gibɛ*	*gbɛ*	'bear'
kul[a] (M)	*kina*	*kul(a)*	*kiɻo(g)*	*kiḷɔg*	*gikil, kilɔ*	*kilɔ*	*ŋku*	'tiger'
mɛt/ḍ	*mʔɔd, amad*	*mad*	*moḍ*	*ɔmɔr/d*	*mɔd, mɔʔ*	*mɔd*	*m-mwaʔ*	'eye'
dʒaŋga	*dʒʔeŋ*	*dʒiʔiŋ*	*-dʒuŋ	*idʒiɲ/ŋ*	*susuŋ*	*tiksuŋ*	*ntʃo*	'foot'
ti ~ tii	*siʔi*	*siʔi*	*tiʔ*	*iti*	*titi*	*titi*	*tti, nti*	'hand'

*Kharia preserves the old stem in *gu**dʒuŋ*** 'wash someone's feet'

Table 1-1: Selected Munda Noun correspondence sets (Anderson 2015b)

Kherwari	Sora	Gorum	Kharia	Juang	Gutob	Remo	Gtaʔ	gloss
-n-	*-n-*	*-n-*	*-n-*	*-n-*	*-n-*	*-n-*	*-n-*	'broom'
RDPL	RDPL	RDPL	RDPL	RDPL	RDPL	RDPL	RDPL	'turmeric'
√¹-a	**kən-√³*	**kən-√³/ *kV-√³*	*√¹-e/-ai*	*√¹-ae*	**kV-√¹*	**kV-√²*	**kV-√²*	'bear'
√²-a	*√¹-a*	*√²-a*	*√¹-ɔg*	*√¹-ɔg*	**kV-√¹, √¹-ɔ*	*√¹-ɔ*	N-√²	'tiger'
Ø	*-ʔ-, *N-*	Ø	Ø	*N-	Ø	Ø	N-	'eye'
-a	*-ʔ-*	*-ʔ-*	--	*N-	RDPL-√	*tik-√*	N-	'foot'
Ø, ˇ	*-ʔ-*	*-ʔ-*	*-ʔ*	*N-	RDPL	RDPL	RDPL, N-	'hand'

Table 1-2: Word formation processes in Table 1

It turns out that Munda languages are not unique in Austroasiatic in showing consistent cognates in the (often CVC) roots of nouns, but some considerable variation in the derivational means to form the syntactically free forms of nouns, often using different prefixes as in the following Katuic (10)-(11) and Palaungic (12) sets.

(10) **Katuic** languages

Bru	Kui	Pakoh	Katu	gloss	Bru	Kui	Pakoh	Katu
kəcah	*kəcah-cah*	*kucah*	*kəcah*	'charcoal'	*kə-*	*kə-*	*ku-*	*kə-*
ncäj	*ncè:*	*nceːᵀ*	*ncaj*	'body lice'	*n-*	*n-*	*n-*	*n-*
ʔəhaːm	*ŋhaːm*	*ʔəhaːm*	*ʔəhaːm*	'blood'	*ʔə*	*ŋ-*	*ʔə -*	*ʔə -*
nluaŋᴮ	-------	*kluaŋ*	*cəluŋ*	'calf, leg'	*n-*	--	*k-*	*cə-*
nciːʔ	-------	*kəciːk*	----	'comb'	*n-*	--	*kə-*	--
ʔəcaː	*caː-ʔacaː*	*ʔəcɔː*	*ʔəcɔ*	'dog'	*ʔə*	*[ʔa-]*	*ʔə -*	*ʔə -*
ruajᴮ	*ruajᴮ-ʔaːruajᴮ*	*rirɔːj*	*rərɔːj* (AD)	'fly'	*Ø*	*[ʔa-]*	RDPL	RDPL
ʔətʌj	*teː-ʔaːte…*	*ʔəti*	*tʌj*	'hand'	*ʔə -*	*[ʔa-]*	*ʔə -*	*Ø*
--	*kətaːm-taːm*	*ʔətaːm*	*ʔətaːm*	'crab' (VN *dam*)	-	*[kə-]*	*ʔə -*	*ʔə -*
ʔəjəːŋ	*ʤiːŋ*	*ʔjiːŋ*	*juŋ*	'foot, leg'	*ʔə-*	*Ø*	*ʔ-*	*Ø*

(Peiros 1996: 1, 2, 4, 7, 10, 28, 57, 67, 69)

(11) So [**Katuic**] (Miller and Miller 1996: 269)
*ra**pay*** 'woman' vs ***mpay*** 'wife' cf. Suai *ka**pay*** 'woman'

(12) **Palaungic** languages (Paulsen 1992: 210-212)

Kontoi	Shinman	Samtao	gloss	Kontoi	Shinman	Samtao
kətam¹	*kaʔ⁴ tam¹*	*tam¹*	'crab'	*kə-*	*kaʔ*	*Ø*
kɔnel¹	*ɛh¹*	*kɔnia²*	'chicken'	*kɔn-*	*Ø*	*kɔn-*
faʔ¹	---	*kɔnfaʔ¹*	'gibbon'	*Ø*	--	*kɔn-*
kɔnkaŋ²	*kaŋ³*	*kɔnkaŋ²*	'rat'	*kɔn-*	*Ø*	*kɔn-*
kɔnvay²	*kaʔ⁴ vai³*	*avai²*	'tiger'	*kɔn-*	*kaʔ*	*a-*
amhạc¹	*kaʔ⁴ mɯik¹*	*mɤc¹*	'ant'	*a-*	*kaʔ*	*Ø*
akʰrak¹	*qhak¹*	*krak¹*	'buffalo'	*a-*	*Ø*	*Ø*
ntak¹	*kaʔ⁴ tak¹*	*n̥tak¹*	'tongue'	*n-*	*kaʔ*	*n-*

Similar such prefix/derivational discrepancies can be found when examining cognate sets in all the Austroasiatic branches, e.g., Vietic (13), Monic (14), or Bahnaric (15), or Nicobarese (16).

(13) **Vietic** Correspondences (Hayes 1992: 222; Ferlus 1974: 73)

Vietnamese	Muong	Ruc	Thavung[1]	gloss	Vietnamese	Muong	Ruc	Thavung
tóc	*thăc*[1]	*ŭsúk*[1]	*sɔk*[1]	---	*Ø*	*Ø*	**n-*	*Ø*
răng	*thaŋ*[1]	*kàsăŋ*[1]	*ksaŋ*[1]‘	*‘tooth’*	**rV-*	*Ø*	**kV-*	**kV-*
---	*lu*[2]	*pŭlu*[2]	*malùu*[2]	*‘thigh’*	--	*Ø*	**pV-*	**pV-* *?/*mV-*
lá	*lá*	*ŭla*[1]	*sláa*	*‘leaf’*	*Ø*	*Ø*	**n-*	**s(V)-*
gà	*ca*[1]	*răka*[1]	*kaa*	‘chicken’	**rV-?*	*Ø*	**rV-*	*Ø*
mưa	*mưa*	*kŭma*[1]	*malòoy*	*‘rain’*	*Ø*	*Ø*	**kV-*	**mV-?*

(14) **Monic** correspondences (Miller and Miller 1996)

Burmese Mon	Thai Mon	Nyah Kur	gloss	Burmese Mon	Thai Mon	Nyah Kur
nìh	*nìh*	***mə**nìh*	‘human’	*Ø*	*Ø*	*mə-*
ŋèak	*ŋìak*	*ŋìak*	‘tooth’	*Ø*	*Ø*	*Ø*
dɒp	*dap*	***kə**dəp*	‘head’	*Ø*	*Ø*	*kə-*
***ka**to*	***ʔa**to*	***kə**tuas*	‘ear’	*ka-*	*ʔa-*	*kə-*
***ka**mot*	***ka**mot*	***kə**mat*	‘fire’	*ka-*	*ka-*	*kə-*
***ka**tac*	***ʔa**tac*	***n**taak*	‘tongue’	*ka-*	*ʔa-*	*kə-*
***pa**soa*	***pa**soa*	***ph**chay*	‘iron’	*pa-*	*pa-*	*p(h)-*
***pa**sa*	***ʔa**sɔ*	*chuun*	‘five’	*pa-*	*ʔa-*	*Ø*
***ta**kah*	***ʔa**kah*	***ŋ**kaa*	‘sky’	*ta-*	*ʔa-*	*n-*

(15) **Bahnaric** correspondences (Bahnar- Jəlәng correspondences) (Léger 1974: 124-5)

Bahnar	Jəlәng	gloss	Bahnar Presyllable/Prefix	Jəlәng Presyllable/ Prefix
anah	*tənah*	'wood, tree'	*a-*	*tə-*
kəja:	*rəja*	‘ginger'	*kə-*	*rə-*
rənga:	*rənga:*	‘sesame’	*rə-*	*rə-*
təmo:	*təmo:*	‘stone’	*tə-*	*tə-*

(16) **Nicobaric** words for 'hand' (or 'palm') (Man 1975 [1888-9])

Central	Car	Shom Pen	Teressa	gloss
*kane-**tai***	*el-**ti:***	*noai-**ti:***	*mòh-**ti:***	'(palm of) hand'

Pan-Austroasiatic correspondences very frequently involve sets of this very type, with a range of different derivational processes and prefixes, as in the following set:

(17) **pan-Austroasiatic** 'earth, land, soil' (Peiros 1998: 251: Field Notes)

Bahnar	Chrau	Mon	Semai	Wa	Deang	Ksinmul	Khmu	Ho	Kui	Khasi
tɛh	*n-tɛh*	*ti*	*te:ʔ*	*tɛʔ*	*k'-tai*	*kə-te:*	*pə-te*	*o-te*	*kə-ta:ʔᴸ*	*pyr-thei*

Prefixes found in this set include **(ʔ)n-; *kə-; o-; *pə-, pər-; *Ø [±*h/ʔ];*

This same situation with cognate roots but non-cognate word-frames is also largely the case when looking at correspondence sets across, for example, the Tibeto-Burman languages of Arunachal Pradesh

(18) Tibeto-Burman languages of Arunachal Pradesh

Bangru	Puroik	Bugun	Sherdukpen	Koro Aka	gloss (Field Notes)
we-se	*sue*	*ga-tɕej*	*zi*	*tu-sũ*	'urine'

Milang	Miju	Idu	gloss
(Tayeng 1976: 17)	(Dasgupta 1977: 74)	(Pulu 2002: 41)	
a-te (a-tə)	*ti-sit*	*the-ci*	'urine'

This type of variation of a shared root embedded within different derivational/compounding structures in cognates can be seen in different local village varieties of Sartang even.

(19) Sartang varieties

a. Khoina Sartang	b. Jerigaon Sartang	c. Rahung Sartang	gloss
*na-**ni***	*ni-**ni***	***ni-**mi***	'sun' (Field Notes)

Indeed this type of variation has been known in Tibeto-Burman studies for some time and has been institutionalized or conventionalized in the

variant reconstructions such as Proto-Tibeto-Burman **s/g-la* 'moon', **d/s-ləy* 'bow', **l/b-ŋa* '5', **g/r-nyi:t* 'sleep' (Matisoff 2003b: 599, 600, 605).

Back to Austroasiatic, neutralization of three distinct Song prefixes to a single one in Chong seems to have occurred in the history of the Pearic branch of Austroasiatic.

(20) **Pearic** Languages (Diffloth 1989: 149)

Song	Chong	gloss	Presyllable	Presyllable
khla̲'a	*kəla̲ʔ*	'leaf'	*k(h)-*	*kə-*
səla̲:	*kəla̲ʔ*	'thorn'	*sə-*	*kə-*
ləpha̲:	*kəpha̲:*	'tortoise'	*lə-*	*kə-*

Pakanic Bugan, in particular, has an intriguing number of recurrent prefixes in these nouns. So many indeed that one wonders if this reflects lexicalization of an earlier semantically transparent system, even if many common ones seem to occur across words that are difficult to unite semantically. So while *pə-* occurs in a number of body parts, it is in far from all and occurs with other words too. So too is *mə-* found in the names of some tools, but also found in words that clearly this meaning has nothing to do with too (21).

(21) Bugan (Li 1996: 137-139, 141, 144)

pə^{55}lai^{33} 'tongue'	*mə^{33}dou^{33}* 'firewood knife'	*tə^{0}qou^{35}*'palm of hand'
pə^{0}mɑ̃^{31} 'nose'	*mə^{33}tsa^{31}* 'hand straw cutter'	*tə^{0}kɔ^{33}* 'tiger'
pə^{0}kui^{55} 'craftsman'	*mə^{0}kua^{31}* 'folk song'	

Some 'classes' however reveal themselves to be possible in Bugan as well, although, again, not all words with these prefixes show this same semantic classification, at least not synchronically or transparently (e.g., 'peach tree' below with what appears to be a female animal classifier).

(22) Bugan (Li 1996: 137-139, 141, 144)

pu- male animals	*mu-* female animals	*tse-* birds
pu^{55}laŋ^{31}'stallion'	*mu^{33}laŋ^{31}* 'mare'	*tse^{0}qa^{35}* 'duck'
	mu^{33}tsau^{33} 'bitch'	*tse^{0}ŋaŋ^{35}* 'goose'
	mu^{33}paŋ^{31}'peach tree'?	

Pakanic Bolyu too has interesting sets of recurrent forms, for example with the prefix *mɔ-*

(23) Bolyu (Edmondson 1995: 139-141)

mɔ-ljaːŋ	*mɔ-fuːn*	*mɔ-tən*	*mɔ-mau*	*mɔ-ti*	cf.	*ɣəm-ti*
'eagle'	'bamboo'	'melonseed'	'finger'	'finger'		'arm'

In the case of the last two, we clearly have a type of word family with the pan-AA root *ti* 'hand, arm' and two different prefixes that specify the meaning. Such word families can be found in other branches as well, for example, the following forms from Aslian Kensiw and Car Nicobarese.

(24) Kensiw (Bishop and Peterson 1994: 176, 178, 185)

tə-kɔʔ 'pus, sputum'
ma-kɔʔ 'egg'

(25) Car Nicobarese (Das 1977: 32)

ɛl-tiː 'palm of the hand'
uk-tiː 'back of the hand'
kun-tiː 'finger'

A number of word families are suggested by different noun forms from Car Nicobarese. Common recurrent initial elements include *ta-, li-,ɛl-*

(26) Car Nicobarese (Das 1977: 31-32, 41, 42)

tarul 'cloud'
tacam 'dew'
tahɯi 'today'
takɯn 'thigh'

li-tak 'tongue'
li-kuɯn 'nape'

(27) Car Nicobarese (Das 1977: 17)

ɛl-ŋoh 'chest'
ɛl-waŋ 'mouth'
ɛl-kui 'brain'
ɛl-mɛh ~ *mɛh* 'nose'
ɛl-ran 'sole, hoof'
but *ɛl-mat* 'color'

Returning to Munda, what we find is that the proto-Munda verb stem could be augmented by a voice prefix or infix, both reflecting an inheritance from Proto-Austroasiatic (Anderson 2004), as well as a combining form of a noun root (or a second verb stem in a serialization/compounding structure, not discussed here but also old). The combining form is cognate across many of the Munda languages, even while the corresponding syntactically free form of the noun in the same languages is not.

(28) Proto-Munda *-*ti* 'hand' (Anderson 2007a: 200)

Remo	Kharia	Juang	Sora	Plains Gtaʔ
*gui-**ti***	*guc-**te***	*guc-**ti***	*leːm-**si**-t-am*	*mbœʔsiʔ gweʔ-**ti**=ke*
wash-hand	wash-hand	wash-hand	bow-hand-NPST-2	left.hand wash-hand=RLS
'wash hands'	'wash hands'	'wash hands'	'I bow to your hands'	'(I) wash(ed) my left hand'

Indeed a cognate system is preserved in the lexicon of numerous Austroasiatic languages suggesting the Munda forms *best* continue a system present in the ancestral Austroasiatic proto-language. Forms of the VN structure can be found in languages representing as diverse group of Austroasiatic as the Vietic, Monic, Aslian, and Pakanic (Mangic) branches.

(29) Thavung-So [Vietic, Thailand]

*khɛʔ**ɛʔ*** 'to shit' (Premsrirat 1996: 168)

(30) Old Mon [Monic] (Nai Pam Hla 1976: 907)

*ti**tey** /ti**tea**/* 'lead' (cf. *tey /tea/* 'hand') cf. modern Mon *_datay /hetoa/*

(31) Kensiw [Aslian] (Bishop and Peterson 1994: 188, 193)

*ka**wɔl*** 'hug' cf. ***wɔl*** 'shoulder blade' *puc**pɛh*** 'swing arms' cf. *k[ə]la**pɛh***'upper arm'

(30) Bolyu [Pakanic (Mangic)] (Edmondson 1995: 134, 141, 144, 154)

*tse**lei***	*vuŋ**qɔ***	*ɬjit**lei***	*ɬjit**tsu***	*ɬjit**tən***
'beat cow'	'to (catch) fish'	'kill cow'	'kill dog'	'butcher pig'

Noun correspondence sets are complex across and within Austroasiatic branches; some of the issues may reflect retention of an earlier now lost system of noun class prefixation. When word frames are non-cognate, the recurrent element–the root–that is cognate can combine with verbs in a system of noun incorporation best preserved in some southern Munda languages like Sora but also found in various other Austroasiatic branches as well and thus was a likely feature of the proto-language.

3.3 Preverbal Negation

Negation is preverbal in the vast majority of Munda languages. Some have innovated the use of negative copula forms into new finite structures and in such cases, negation is post-verbal, as negative copula forms appear in the clause and usually in sentence final position as well. Indeed, both are found in the Juang example in (32) *-ma-* in pre-verbal position and *ɟena* in sentence-final position.[2]

(32) Juang (Patnaik 2008: 546)

apa	*a-*	*ma -*	*ɟim -ke*	*ete*	*aiɲ*	*kikib*	*ɟena*
2DL	2DL-NEG-	eat-PRS	because	I		RDPL~do	NEG.COP

'Because you don't eat (it), I didn't do it' preverbal/postverbal

The *ma*-negator of Juang is cognate with the non-finite and attributive negator of Gtaʔ, where it is realized as *ma-*.

(33) Gtaʔ

ma	*=bihæ =nə*	*ngire*
NEG.ATTR=	marry=ATTR	young.man

'Unmarried young man, bachelor' preverbal

Possibly connected to this is the preverbal negator of Kharia *um*. Note that the morphosyntax of this element in Kharia favours subject clitics on the negator–a pattern almost certainly calqued on Kherwarian, specifically Mundari, patterns in Kharia (Anderson and Jora 2018), where the pattern is cognate but not the negator involved (which is *ka* in Mundari).

(34) Kharia (Peterson 2008: 463)

um=iɲ *ter=e*
NEG-1 give=IRR
'I won't give' preverbal+SUBJ

(35) Tamaɽia Mundari

aĩja(ʔ) *ti* *ka=iŋ* *abuŋ-a*
I:GEN hand NEG=1SG.SUBJ wash-IND
'I will not wash my hand' preverbal+SUBJ

These pre-verbal *m*-negators appear to have analogs in the eastern branches of Austroasiatic as well, including a diverse array of groups as Khasian, Bahnaric, and Mangic.

(36) Mangic Bugan (Li & Luo 2015: 1042) *mə* NEG preverbal

(37) Bahnaric Bunong (Butler 2015: 739) *mo:* NEG preverbal

(38) Standard Khasi (Nagaraja 2015a: 1177) *=m* NEG preverbal

NB: host to preverbal clitics! *ka=m* 3FEM=NEG *u=m* 3MASC=NEG *ki=m* 3PL=NEG

Other negators in Munda also tend to be in preverbal position. This includes the Kherwarian prohibitive particle *alo*, the first part of which may be cognate with the bound negator *a(r)-* in languages like Sora or Hill Gtaʔ.

(39) a. Birhor b. Birhor

***a**lo=m* *nir=a* ***a**lo=m* *gitiʧ=a*
PHB=2SUBJ run-IND PHB=2SUBJ sleep-IND
'Don't run!' 'Don't sleep! Preverbal+subj

(40) Sora (Anderson & Harrison 2008b: 346)

ɲen *bazar-ɪn* *ə-je:r-ej*
I market-N.SFX NEG-go-1
'I don't, won't go to the market' preverbal

(41) Hill Gtaʔ

kine hãwe a-na n-a-biʔ
this bow OBJ-you 1-NEG-give
'I will not give you this bow'

(42) Hill Gtaʔ

kine hãwe a-na m-biʔ=wɛ
this bow OBJ-you 1-give=IRR/FUT
'I will give you this bow' preverbal

Negative formations in Austroasiatic languages are typically many in number, but preverbal negative scope operators are the rule in all constructions except very recently innovated extensions of negative copula forms into finite inflection in certain endangered Munda languages like Juang or Gutob.

3.4 NP Order and Classifiers

I limit myself here to the relative order of the noun (N) to the numeral (NUM) and classifiers (CLF) that typify most Austroasiatic languages and which fall into two large patterns. Otherwise, I will only discuss very simple NP structures and restrict the discussion to the relative order of the N to a possessor (GEN) to adjectives (ADJ) and demonstratives (DEM) and classifiers (CLF).

With respect to the tripartite collocations of quantified NPs with numerals, two broad patterns are found in Austroasiatic. One pattern we can broadly describe as the Taic pattern has the Noun in phrase-initial position followed by the numeral and then the classifier, i.e. N NUM CLF. This is also characteristic of what might be called the core of the Austroasiatic region or its centre, typifying as it does both where Austroasiatic languages overlap with the core Taic-speaking areas. So in addition to Khmer and Mon, this pattern is found in Pearic, Khmuʔic, Palaungic, and Mangic, as well as in Khmericized/Taicized Bahnaric and Katuic languages like Bunong or Kui Ntua.

The second broad pattern could be broadly characterized as Sinitic or Sino-Viet. This pattern has the noun at the end of the quantified phrase, with the numeral preceding the classifier and both preceding the noun, i.e., NUM CLF N. This characterizes what could be called the peripheral Austroasiatic region, typifying as it does the eastern and western branches of the family, occurring in Vietic, less restructured Katuic and Bahnaric languages, Khasic, Nicobarese, and even Munda.

The relative order of the other elements within the noun phrase follows some broad areal patterns, but the earlier system seems relatively straightforward to reconstruct. Head-initial NP structure typifies all of the Southeast Asian branches of Austroasiatic. Leaving aside the non-trivial

issue for the present about whether adjectives really exist as a part of speech or word class in Austroasiatic languages (for example, they appear simply to be a type of stative verb in Old Mon, Old Khmer, and Khasic languages as a whole), the order N ADJ is found in all Southeast Asian branches of Austroasiatic. Similarly, all such branches attest to basic N DEM and N GEN order. Basically, there is a cline towards South Asian head-final structure within NP-internal syntax seen when comparing Khasic, Nicobarese, and Munda with the branches that are located firmly within the Mainland Southeast Asia linguistic area. Munda unsurprisingly shows the greatest degree of restructuring on South Asian linguistic models, with DEM N, GEN N, and ADJ N, but there is internal evidence in Munda that suggests that N GEN and N ADJ order were found at one point. Nicobarese shows ADJ N order but maintains the older N GEN order. Both Nicobarese and Khasic as well as Munda, show the DEM N order that typifies South Asian languages more broadly. These are summarized in Table-3.

Vietic	NUM CLF N	N DEM	N GEN	N ADJ
Bahnaric*	NUM CLF N	N DEM	N GEN	N ADJ
Katuic*	NUM CLF N	N DEM	N GEN	N ADJ
Pearic	N NUM CLF	N DEM	N GEN	N ADJ
Khmeric	N NUM CLF	N DEM	N GEN	N ADJ
Mangic	N NUM CLF	N DEM	N GEN	N ADJ
Khmuʔic	N NUM CLF	N DEM	N GEN	N ADJ
Monic	N NUM CLF	N DEM	N GEN	N ADJ
Palaungic	N NUM CLF	N DEM	N GEN	N ADJ
Aslian	NUM CLF N	N DEM	N GEN	N ADJ
Khasic	NUM CLF N	DEM N	N GEN	N ADJ
Nicobarese	NUM CLF N	DEM N	N GEN	ADJ N
Munda	NUM CLF N	DEM N	GEN N	ADJ N

Table 1-3: NP-internal syntax in Austroasiatic languages
* some Khmericized/Taicized Bahnaric and Katuic languages have N NUM CLF

This distribution would suggest the following structures for Proto-Austroasiatic:

(43) Proto-Austroasiatic NP-internal syntax

NUM CLF N N DEM N GEN N ADJ

So the original structures would then be preserved in Vietic and Aslian, and the less restructured varieties of Bahnaric and Katuic. Munda has moved closest to the South Asian NP syntax having reversed the order of GEN and N, while Khasic and Nicobarese have both also shifted to DEM N order and Nicobarese also to ADJ N order. So we see a cline of accrual of South Asian syntactic features in the South Asian Austroasiatic branches that shows a cline from least to most restructured of Khasic > Nicobarese > Munda. What all of this means of course is that these are yet further data that show the concept of rhythmic holism entailing one-time parametric resets of Munda from the MSEA to SA areal norms is nonsense, and that different groups in South Asia have accrued such secondary features at different times and in different domains, and that the false dichotomy between Munda and all other AA language branches remains untenable.

4.0 Periodization of Clear South Asian Contact Features in Munda

4.1 SOV order: proto-Munda

Most Austroasiatic languages and branches have dominant verb-medial syntax, with subject/actors/agents typically preceding the verb, with subcategorized objects and adjuncts following. This includes for example the Aslian language Maniq (44), Pearic Chong (45), Khmer (46), Bahnaric Bunong (47), Katuic Kui Ntua (48), Khmuʔic Kammu (49), Mangic Bugan (50), or Standard Khasi (51), Vietnamese (52), or Palaungic Dara'ang Palaung (53).

(44) Maniq

yaʔ	*ʔɛʔ*	*ʔɛk*	*kut*	*ʔey*
woman	3	give	bottle	man

'The woman gave the man the bottle'
(Kruspe et al. 2015: 438)

(45) Chong

dak	*kāmlāŋ*	*tàt*	*jā:ŋ*
3H	PROG	rive	rubber.tree

'He is riving rubber trees (for making a resin)'
(Premsrirat and Rojanakul 2015: 611)

(46) Khmer

khŋom	*ʔaoj*	*luj*	*koət*
I	give	money	3

'I gave him money'
(Bisang 2015: 704)

(47) Bunong

ŋit	*ən*	*tap*	*ta*	*ɲc^{h}ot*
Ngit	give	egg	LOC	Nchot

'Ngit gave an egg to Nchot'
(Butler 2015: 738)

(48) Kui Ntua

na:w	*tuəŋ*	*cuəh.t^{h}a:w*	*pɨ:t*	*ʔɛŋ*
3	fear	tiger	large	ANA.DIST

'They are afraid of that big tiger'
(Bos and Sidwell 2015: 859)

(49) Kammu

mè	*mʌ̀t*	*mɨ́ar*	*ʔòʔ*
2SG.M	take	loincloth	1SG

'Did you take my loincloth?'
(Svantesson and Holmer 2015: 971)

(50) Bugan

li^{55}	*sai^{33}*	*tso~ʦ^{31}*	*tsiu^{55}*	*naŋ^{31}*
ox	DUR	eat	grass	DUR

'The ox is eating grass'
(Li and Luo 2015: 1054)

(51) Standard Khasi

u	*la*	*a:j*	*ja*	*ka=kot*
3M	PST	give	ACC	FEM=book

'He gave the book'
(Nagaraja 2015a: 1172)

(52) Vietnamese

Hương	*cho*	*họ hàng*	*thuốc*
Huong	give	relatives	drug

'Huong gives relatives drugs'
(Brunelle 2015: 925)

(53) Dara'ang Palaung

ʔo	*dɨ*	*măh*	*mɛ*
1SG	GOAL	hit	2SG

'I will hit you'
(Deepadung et al. 2015: 1075)

Curiously, a number of Austroasiatic languages have basic verb-initial order. Included in this group is Car Nicobarese which apparently has dominant verb-initial order in main clauses (Sidwell 2015b). Note however that subordinate clauses have verb-medial order, which suggests that the verb-initial order in main clauses could well be a secondary development under Austronesian (Acehnese?) influence. Nancowry (Muot) on the other hand has a preferred verb-initial order in main clauses too, with VPA or VAP both attested:

(54) Nancowry (Muot)

kalóʔ	*nɔ́t*	*cə́-n*	*kamalóʔ*
steal	pig	my-NOM	theif

'The thief stole my pig'
(Reid 1994: 333)

However, some languages for which no such contact influence can be attributed also show verb-initial dominant order in Austroasiatic. This set includes most of the languages in Khasic other than Standard Khasi itself

and its closely related lects. It does, however, include the Khasic languages Amwi and War (Jenny et al. 2015: 59) as well as Pnar (55).

(55) Pnar

tʃim u=bru ka=waʈ
take MASC=person FEM=word
'The man took the sword'
(Ring 2015: 1196)

While Palaungic languages as a whole are AVP (SVO) languages, including Dara'ang Palaung (10) above, and Danau as well (Si 2015), Wa languages have dominant verb-initial order as well (Jenny et al. 2015: 61), e.g., Tangyan Wa.

(56) Tangyan Wa

sɔm ʔɤʔ ʔɯp
eat I rice
'I eat rice'
(Jenny et al. 2015: 62)

Verb-initial order is now considered to be the original order of Proto-Austroasiatic (Jenny 2015, 2020, Jenny et al. 2015, 2017). Note that this means that significant variation can exist in relatively closely related languages within a sub-phyletic branch in AA, e.g., in Khasian: Standard Khasi SVO (Nagaraja 2015a) vs. Pnar VSO (Ring 2015).

Nevertheless, Munda stands apart from the rest of Austroasiatic with its verb-final order. Munda clause-level constituent order thus remains quite distinct within the phylum. Ho (57), Korku (58), Remo (59), and Plains Gtaʔ (60), etc., are all SOV. Occam's razor compels us to reconstruct what is the simplest solution, and that is to assume a one-time innovation at the Proto-Munda level to verb-final structure. Note that the entire pre-Proto-Munda verbal predicate, with preverbal operators encoding a range of verbal categories already prosodically/morphotactically tied to the verb stem and its post-verbal operators, moved to the final position. Therefore, all Munda languages now have these operators either within the verb or immediately preceding it, but all this in clause-final position relative to NP arguments and adjuncts.

(57) Ho

alaŋ *ɖaʔa* *ka=laŋ* *ɲu-e* *tʃa=laŋ* *ɲu-e*
we.2 water NEG=1DLdrink-IMPtea=1DL drink-IMP
'let's not drink water, let's drink some tea'

(58) Korku

iɲdʒ *dukana:-ʈen* *sa:kar* *sasa:-bà*
I store-ABL sugar RDPL.bring-IND
'I will bring sugar from the shop' (Nagaraja 1999: 71)

(59) Remo

niŋ *nsuɽaʔ* *susum* *ɖen-t-iŋ*
I banana RDPL~eat AUX-NPST-1
'I am eating a banana'

(60) Plains Gtaʔ

kæn *sela-^{M}boeʔ* *huŋ-ɖæ=ke* *bæ=ke*
this.one girl=child/FEM child-3.REF-OBJ send-RLS
'He sent his daughter'

Both Indo-Aryan (Masica 1993) and Dravidian (Krishnamurti 2003) are reconstructed as verb-final. All the Tibeto-Burman languages of South Asia are verb-final (although what Proto-Tibeto-Burman was is controversial, cf. Dryer (2003, 2008) and not really relevant to Proto-Munda most likely). Nihali (Nagaraja 2015b), Burushaski (Anderson 2007b), and Kusunda (Watters 2006) are all verb-final. Thus, all of the language groups that Munda would have ever come into contact with possibly in South Asia share this syntactic feature and thus the shift to verb-final word order in Munda likely reflects a process of accommodation to this as Proto-Munda speakers encountered various population groups over what was most likely the entire Proto-Munda period, but nevertheless pre-dating the breakup of Munda into the attested sub-groups.

Note, however, that while Proto-Munda was SOV, there is both internal and external evidence that this is a secondary feature having originally been VSO with a pragmatically determined variant in SVO that characterized Proto-Austroasiatic (Jenny et al. 2015a, 2015b; Jenny 2016ms).

Internal Munda evidence comes from the verb-noun compounds in Munda.[3]

4.2 Objective case marking in *-ke*: each language individually

The objective case form in *-ke* seen in various Munda languages has both different Indo-Aryan sources and different functions or distributions in the languages concerned. For example, the Kherwarian Munda language Kera? Mundari has *-ke* as a case suffix. Functionally, the element is an accusative-dative marker, which encodes either (61-63) or both (64) of these functions in the same clause, and can appear with both animate and inanimate nouns.

(61) Kera? Mundari

aiŋ-ke *alɔ* *kaj-iŋ=me*
I-OBJ PHB tell-1SG.OBJ=2SG.SUBJ
'don't tell me!' (Field Notes)

(62) Kera? Mundari

am *aiŋ-ke kudaɔ=ki-a=m*
you I-OBJ run:CAUS=TAM.TR-IND=2SG.SUBJ
'you made me run' (Field Notes)

(63) Kera? Mundari

am *aiŋ-ke ka=m* *kudaɔ-t-ĩ-a*
you I-OBJNEG=2SG.SUBJ run:CAUS=TAM-1SG.OBJ-IND
'you didn't make me run' (Field Notes)

(64) Kera? Mundari

aiŋ-ke *muʈri-ke* *alɔ* *em-ku=m*
I-OBJ basket-OBJ PHB give-3PL=2SG.SUBJ
'don't give me the baskets!' (Field Notes)

Note that this case element is a dependent marking strategy that coexists with a head-marking strategy where the referent is encoded in the verb as an object in the verb (61, 63). Note that object agreement is not

obligatory in such formations and the case-marked argument may also not be encoded in the verb as well (62).

In its close sister language Tamaɽia Mundari, however, the segmentally identical element is morphotactically, functionally, and distributionally different. In Tamaɽia Mundari, the case marker is extrametrical and thus functions as a clitic, not as a suffix. Further, it plays a role in a system of differential object marking in Tamaɽia Mundari, such that inanimate objects lack *=ke* on the NP but are still inflected in the verb agreement (65), while animate ones are encoded by both (66)-(69). As in Korku (see below), the case appears with only one argument in the clause at one time and has the function of an objective case marker in Tamaɽia Mundari.

(65) Tamaɽia Mundari

hoɽo	*koto*	*ka=i*	*rapud*	*daɽ-i=a*
man	branch	NEG=3	break	FUT-3-IND

'the man will not break the branch' (Field Notes)

(66) Tamaɽia Mundari

aiŋ	*hon=ke*	*ka=iŋ*	*abuŋ=k-i=a*
I	baby=OBJ	NEG=1	wash=TAM-3-IND

'I did not wash the baby' (Field Notes)

(67) Tamaɽia Mundari

aiŋ	*hon=ke*	*ka=iŋ*	*abuŋ=Ø-ij=a*
I	baby=OBJ	NEG=1	wash-3-IND

'I will not wash the baby' (Field Notes)

(68) Tamaɽia Mundari

kula	*sukri=ke*	*goiʔ-k-i-a*
tiger	pig=OBJ	kill-PFV.TR-3-IND

'the tiger killed the pig' (Field Notes)

(69) Tamaɽia Mundari

kula	*sukri=ke*	*ka= i*	*goiʔ-k-i-a*
tiger	pig=OBJ	NEG=3	kill-TAM-3-IND

'the tiger did not kill the pig' (Field Notes)

Tamaɽia Mundari has a mixed-head and dependent marked system here too. Objects are encoded in the verb form as well, not just by the objective case. According to Osada (1999: 53), *=ke* in Mundari and Ho (71) and *-ke* in Kera? Mundari is a form copy or borrowing from Sadani/Sadri (70) (Nowrangi 1956, Jordan-Hortsmann 1969, Kiran and Peterson 2011, no date).

(70) Sadani/Sadri

Sadi ghʌr=wala=mʌn u chʌgri=ke kaʈ-l-ʌẽ ʌur kha-l-ʌẽ
Wedding house=ADJVZR=PL that goat=OBL cut-PST-3PL and eat-PST-3PL
'The people of the wedding house…cut the goat up and ate it'
Kiran and Peterson (Киран and Петерсон [2011])

Note that with respect to the Kherwarian language Ho, according to Pucilowski (2012: 20), *=ke* "was used often in my elicitation work with young, educated, and (Sadani-/Sadri-/Hindi-) bilingual students...[but it] is not considered grammatical by older speakers." This suggests that its origin is quite recent in Ho and it is not fully integrated into the grammar of the language yet. Note that the older system is the head-marking system in Kherwarian with the object agreement in the verb, as seen in the following Ho example, where *-i-* in the verb indexes the argument *seta=ke*.

(71) Ho

Dobro=do seta=ke hapa-n-me meta-i-ten-e
Dobro=FOC dog=ACC quiet-RFLXV-2SG.IMP say.APPL-3SG-IPFV-IND
'Dobro says to his dog "be quiet!"' (Pucilowksi 2012: 20)

In each of the above three languages, the source of the loan is almost certainly the tribal Indo-Aryan lingua franca of Jharkhand Sadani/Sadri/Nagpuri. However, given the different morphotactics and functional domains exhibited, even when looking at two varieties of Mundari, it appears that each language borrowed the element at different times but from the same loan source. This seems to have happened only very recently in Chaibasa Ho as older speakers still do not use *=ke*.

In North Munda Korku, a sister to the Kherwarian branch, the functional element under discussion is a primary objective case suffix (Dryer 1986) that encodes either a patient in a two-place argument structure or recipient in a three-place one.

(72) a. Korku (Mohan 2015: 6)

porija siɲ-ke mama-lakken
boy tree-OBJ cut-PROG
'The boy is cutting the tree'

b. Korku (Mohan 2015: 8)

iɲ diku-ke kitab dʒi-ke
I they-OBJ book give-PFV.TR
'I gave them the book'

In form and function, this Korku *-ke* element likely represents a form copy from the local variety of Indo-Aryan Nimadi, which has an identical element, and thus is not likely to be the same as the Nihali dative/instrumental case form *-ki*, which is both functionally and formally distinct.

In the southern Munda Language Plains Gtaʔ of Odisha, one also finds the use of a case clitic *=ke* on (pro) nominal objects in a still different system of differential object marking (e.g., Sinnemäki 2014). In Plains Gtaʔ, the case element occurs as a primary object marker on NP objects (73).

(73) a. Plains Gtaʔ

meʔ-swa=ne bba-ɽœ huŋ-ɖœ-hiŋ=ke saliaʔ+ku=ke
one-day=GEN father-3.REF child-3.REF-PL=OBJ ask+ask=NFUT
'One day their father asked his children…' (Field Notes)

b. Plains Gtaʔ

ɖokra mriaʔ=ʧe ɖukri=ke basoŋ=ke
old.man rise=SS old.woman=OBJ say=NFUT
'The old man got up and said to the old woman… (Field Notes)

c. Plains Gtaʔ (Field Notes)

at=baŋ eʔke=n(e) remwa-hiŋ=ɖig hliʔ=ke
there=ABLtoday=GEN person-PL=also bamboo.shoot=OBJ
ɖwe-ʧe ʧʧoŋ-miaʔ-har=ke
cook-SS RDPL~eat-HAB-PL=NFUT
'From that time on the people now a days cook and eat bamboo shoots'

With pronominal objects, on the other hand, *=ke* marked pronouns also require the older etymological objective case prefix *a-* to be included as well (74). For more on this case marker, see below.

(74) a. Plains Gtaʔ

a-nœ=ke *kmœ-hiŋ* *a-miaʔ*
OBJ-we=OBJ DEF3-PL NEG-know
'they don't know us' (Field Notes)

b. Plains Gtaʔ (Field Notes)

nœŋ hare=ge a-nœŋ=ke siʔ-har=ke ɖaktʃe basoŋ=ke
I defeat=EVID OBJ-I=OBJ cut-PLURACT=NFUT QUOTsay=NFUT
'"I am defeated it seems; they have hacked me off," he said'

Plains Gtaʔ also marks 'dative' experiencer 'subjects' with certain predicates with the *=ke* case marker as well.

(75) Plains Gtaʔ (Field Notes)

tæn kitoŋ remwa=ke bweʔtur=la, remwa=ke gœ=ke
that god person-OBJ spit=DS person-OBJ itch=NFUT
'That god spat on the person, and the person started itching'

In closely related Hill Gtaʔ, the form-copied case clitic of Indo-Aryan origin has been integrated into the language in a different system of differential object marking: here, *nominal* NP objects are marked by *=kə* (76), but *pronominal* objects are marked only by *a-* (77). Note that no object encoding is attested in the verb in Plains or Hill Gtaʔ, so it is a purely dependent marked system.

(76) Hill Gtaʔ

ŋku gubuʔ=kə goʔ=gə
tiger pig=OBJ die=EVID
'The tiger killed the pig' (Field Notes)

(77) a. Hill Gtaʔ

a-nœjŋ *basoŋ*
OBJ-I tell
'Tell me!'

b. Hill Gtaʔ

a-nœjŋ *basoŋ-pe*
OBJ-I say-2PL
'Tell me y'all!' (Field Notes)

c. Hill Gtaʔ

a-nœjŋ ɖali biʔ-la
OBJ-I basket give-IMP.
'Give me the basket!'

d. Hill Gtaʔ

a-nœjŋ ɖali biʔ-la-pa
TROBJ-I basket give-IMP.TR-2DL
'Give me the basket you 2!'

e. Hill Gtaʔ

a-nœjŋ ɖali biʔ-la-pe
OBJ-I basket give-IMP.TR-2PL
'Give me the basket y'all!'(Field Notes)

The likely source for the object case clitic *=ke* in Plains and Hill Gtaʔ is Desia, which also has a dative/accusative or primary object marker in *-ke* (Datta 2002: 99). Desia is the locally dominant tribal Indo-Aryan variety, which both endangered Gtaʔ varieties are shifting to presently. Note that this case element *-ke* is found in many of the non-standard and 'tribal' varieties of Odia (Koul 2002: 180-190) and Bhatri (Ghosh 2002: 143).[4]

Given *-ke/=ke* in a dative-accusative or objective function in Plains Gtaʔ, partially in Hill Gtaʔ, as well as in Tamaɽia Mundari, Keraʔ Mundari and in Korku, might tempt one to assume that this is an archaic feature to be reconstructed back to Proto-Munda. However, the case element occurs in precisely those Indo-Aryan languages that dominate and stand in an asymmetrical bilingual relation with the Munda languages that show it and moreover other Munda languages lacking this particular contact milieu and even some that share it show no traces of this and the morphosyntax of object encoding is rather different. For example, Kherwarian Santali lacks any evidence for *=ke.* Only object agreement morphosyntactically encoded in the verb form, and the preferred pre-verbal (or pre-negator) syntactic position demarcates objects in Santali (78).

(78) a. Santali

am iŋ=em ɖaɽ-otʃo-ki-d-iŋ-a
you I=2SG.SUBJ run-CAUS-PFV.TR-TR/ACT-1SG.OBJ-IND
'You made me run' (Field Notes)

b. Santali

am iŋ ba=m ɖaɽ-otʃo=li-d-iŋ-a ~ ɖaɽ-otʃo=ki-d-iŋ-a
you I NEG=2 run-CAUS-TAM.III-TR-1-IND run-CAUS=TAM.TR-TR-1-IND
'You didn't make me run' (Field Notes)

The case marker *=ke* is also not used at all in Remo either. Remo uses an objective case proclitic/prefix *a-* with both pronouns (79) and nouns (80) alike, and thus shows no split in the manner of Plains Gtaʔ or Hill Gtaʔ. In Remo, indefinite inanimate nouns (81) tend not to be marked by the case element. Note that the case proclitic can attach to attributive adjectives (82) not just the (pro) noun in Remo, so the element has different morphotactics and syntactic distribution and operates within a different type of differential object marking system than in Gtaʔ.

(79) Remo

niŋ a-no dʒuʔ-t-iŋ
I OBJ-you see-NPST-1
'I see you' (Field Notes)

(80) Remo (Fernandez 1968: 66)

niŋ a-remo kijaŋ bibe(ɖ)=ɖen-t-iŋ
I OBJ-man rice RDPL~give=AUX-NPST-1
'I am giving rice to the man'

(81) Remo

niŋ nsuɽaʔ susum ɖen-t-iŋ
I banana RDPL~eat AUX-NPST-1
'I am eating a banana' (Field Notes)

(82) Remo (Fernandez 1968: 119)

gitin remo a-monaʔbaj selane kijaŋ beɖ-oʔ
DEM man OBJ-fat girl rice give-PST.I
'That man gave the rice to the fat girl'

Thus, the apparent correspondences in the Munda languages with *-ke/=ke* seem to be just that, i.e., they represent parallel developments in

the languages where such case markers have been borrowed from different IA sources. This is in part supported by how each language that has copied this form shows different morphotactics and different functional distributions and systems of oppositions the form has been integrated within.

What *might* be old in Munda however is the case prefix. Within Munda, there is evidence that we might want to reconstruct this element as a dative/objective case marker appearing on (pro) nouns. These appear as both a proclitic or prefix on the NP in southern Munda languages like Remo or Gtaʔ, and both as unexplained forms of freestanding pronouns in some but not all Kherwarian languages, and incorporated in the verb reinterpreted as an applicative marker in Proto-North Munda, seen in the following Ho vs Hill Gtaʔ examples.

(83) Ho

aiŋ	*am*	*ka=iŋ*	*nel-**a**-me-a*
I	you	NEG=1SUBJ	see-APPL-2OBJ-IND

'I did not look at you'

(84) Hill Gtaʔ

nœjŋ	***a**-na*	*n-a-kej=tə*
I	**OBJ**-you	1-NEG-see=NEG.PST

'I did not see you'

Munda is not alone within Austroasiatic showing possible reflexes of this. Thus, Alves (2004, 2015) mentions dative forms of personal pronouns that take a prefix *ʔa-* in Pacoh. In form and function, this Pacoh (85) appears identical to a subset of contexts of use that characterize the Gtaʔ formation.

(85) Pacoh (Alves 2015: 889)

kɨ:	*pacɔ:m*	*ʔa-maj*	*kaŋ*	*ʔaŋ*
1SG	teach	DAT-2SG	language	English

'I teach you English'

In this discussion, it is important to understand that the degree of morphological integration with a head of semantic operators has nothing to do with whether the elements concerned are grammaticalized. Also, there are meta-analytical filters operating that favour certain types of analysis

over others *a priori* in Austroasiatic languages of Mainland Southeast Asia (and alas beyond where such a belief is less justified). One such example is the belief that Southeast Asian languages are isolating and thus lack inflection, and indeed any morphological processes except compounding and some very limited derivation; some examples demonstrating the fallacy of this belief were offered in 3.1 above. Even in descriptions of languages that clearly have considerable morphology like Car Nicobarese, the operation of this meta-analytic filter is at work, and extracting this information from Sidwell (2015b), for example, often requires some considerable effort into unravelling the glosses to determine their internal morphemic structures. Thus, I consider bound prefixes or proclitics (or suffixes/enclitics) used in a 'normal' case marking pattern as well as grammaticalized case 'prepositions' such as the accusative and dative prepositions in Standard Khasi to be inflections since while they target phrases and are thus not prefixes, morphotactically they do appear to have to appear immediately before the N(P) they govern and thus appear to be proclitics. Since clitics are often but one stage in the cline to full morphotactic integration to an affix, I consider clitics that are functional operators to be just as viable an index of inflectional categories as ones that are affixal. In other words, the degree of morphotactic integration is a separate phono-prosodic parameter than semantic functional grammaticalization. If either case element in Khasi were potentially cognate with Pacoh or Munda, the dative in *ha* would be the more likely candidate, but both cases in (86) are certainly grammaticalized elements in Khasi nevertheless.

(86) Standard Khasi

u	*la*	*a:j*	*ja*	*ka=kot*	*ha*	*ŋa*
3SG.M	PST	give	ACC	FEM=book	DAT	1SG

'he gave the book to me' (Nagaraja 2015a: 1170)

This may well be cognate with the locative *haʔ* in Aslian languages like Semaq Beri, which is functionally similar to the Khasi form.

(87) i. Semaq Beri (Kruspe 2015: 507)

ʔəŋ	*knãl*	*haʔ*	*ja*
1SG	know	LOC	2SG.F

'I know you'

ii. Semaq Beri (Kruspe 2015: 507)

kɛ	*nɔ~t*	*haʔʔəŋ*
3SGsee	LOC	1SG

'she looked at me'

As a verb-initial language, Proto-Austroasiatic would have likely innovated such functional operators from a serial verb construction and thus the development of something like what has happened in Modern Mon in the shift of *kɒ* 'give' to an adposition meaning 'to' (Jenny 2015: 586). The development of the object case markers in the various Austroasiatic branches, including Munda suggests a similar derivation.

5.0 Referent Indexing in Munda does not reflect Contemporary South Asian Contact History

The Proto-Munda finite declarative verb template, at least one instantiation of it, took the form of a verb stem with some internal structure possibly preceded by a clitic/affix chain of up to three functional operators and followed by a clitic/affix chain of up to three. Causative and reciprocal were variably a prefix or an infix, depending on whether the verb stem was light (simplex) or heavy (complex), but other voice/valence categories were expressed by suffixes. What is important to note here is that the verb in Proto-Munda likely agreed with both subjects and objects, see examples from Juang (89) and Gorum (90).

(88) *SUBJ=NEG=<AM_I>[$VOICE_I$=]Verb.Stem=TAM_J:$VOICE_J$(=OBJ)

(89) a. Juang

mɛ-dʒɔ-ki-ɲ
2-see-PRS.TR-1
'You see me'
(Matson 1964: 35)

b. Juang

e-dʒo-e-neniŋ
2PL-see- FUT.TR-1PL
'y'all will see us'
(Matson 1964: 35)

(90) a. Gorum (Aze 1973: 249-50)

mo-taʔj-iŋ
2-give-1
'You gave me'

b. Gorum

ne-aʔj-t-om
1-splash-NPST–2
'I will splash you'

It is not clear what the clausal alignment of Proto-Munda was. Most show accusative patterning but Sora and Juray (91)-(96) also show a semantic sensitivity (Anderson and Gomango 2017), so the issue remains open.

(91) Juray

əman ənin=adoʔŋ batoŋ-t-am
you 3.PRON=OBJ fear=N.PST-2UND
'You are scared of her'(field notes)

(92) Juray

ɲen əman giɟ-l-am
I you see-PST-2UND
'I saw you' (field notes)

(93) Juray

ɲen ənin=adoʔŋ batoŋ-l-iɲ
I 3.PRON=OBJ fear-PST-1UND
'I was scared of her' (field notes)

(94) Juray

əman ɲen giɟ-l-iɲ
you I see-PST-1UND
'You saw me' (field notes)

(95) Juray

ɲen ənin=adoʔŋ kan-əti tij-t-ai
I 3PRON=OBJ this give-NSPT-1ACT
'I will give this to her' (field notes)

(96) Juray

ɲen je-t-ai
I cry-NPST-1ACT
'I (will) cry'(field notes)

In languages with accusative alignment, there are Munda languages with a primary object patterning (Dryer 1986) like Gtaʔ, Gorum, and Mundari and ones with an accusative object patterning like Korku and differently, Santali, so again, the specific details of the functional oppositions active in the proto-language system remain open questions at present.

Object agreement is otherwise unattested in Austroasiatic and appears to be a clear innovation at the proto-Munda level, albeit expressed by a series of formal markers morphotactically more integrated into the verb, specifically into the TAM-markers, than even subject markers might be in individual languages, e.g., Kherwarian. Subject encoding in the verb is also very rare in non-Munda Austroasiatic but it is attested in Katuic Pacoh and Aslian Temiar.

(97) Temiar (Benjamin 1976: 175)

yeːʔ ʔi-tɛrsəg cɛp
I1SG-trap bird
'I trapped the bird'

(98) Pacoh (Alves 2004: 39)

ʔi-taʔ pəlloː ʔalɔːŋ
UNSPEC-make tube wood
'One makes a wooden tube'

Other traces of subject marking in Austroasiatic can be found in Aslian Che Wong in Kruspe et al. (2015: 436) in the form of prefix agreement

obligatorily appearing with dynamic verbs and the Khasi gender-cum-person subject proclitics found in NEG & FUT constructions.
Polypersonal verbs are not at all common in Dravidian (Krishnamurti 2003: 307ff.) and restricted in Indo-Aryan to certain languages like Kashmiri, Marathi, or Maithili (Masica 1993: 261) which may well show non-Indo-Aryan substrate or contact effects at least in part, triggering or supporting the development of polypersonal verbs in them. On the other hand, the only real analog to the Munda situation is found in Burushaski of northern Pakistan and in Kiranti languages of Nepal. In both situations, polypersonal verb forms are typical and characteristic.

(99) Burushaski [Isolate; Pakistan]

dʒáa	*a-yúguʃanc*	*moó-y-a*	*bá-a*
I.GEN	1-daughter.PL	2PL-give-1	AUX-1

'I herewith am giving you my daughters' (Berger 1998: 161)

(100) Limbu [Kiranti]

sɛʔr-u-ŋ	*nɛtt-u-ŋ*
kill-3-1.PRET	AUX-3-1.PRET

I was about to kill him' (van Driem 1987: 125)

As all three groups are on the periphery of South Asia now, with Indo-Aryan/Dardic and Dravidian languages having expanded into the core areas, leaving the former population groups marginalized in high-altitude or otherwise less desirable agricultural lands, and the languages of the now peripheral groups share various typological features, it is at least possible that the rise of polypersonal structure in Munda verbs was in part triggered or reinforced by contact with pre-Arayo-Dravidian populations at an early period in South Asia.

Note that it is not only referent indexing but other systems also that show sensitivity to the valence of the predicate in Munda languages. Thus, we find languages like the Mayurbhanj (Odisha) variety of Ho, where the default 'past' TAM marker is the anterior *=le* for one-place predicates or detransitives (101) and the perfective *=ke* for two place ones (102).

(101) Mayurbhanj Ho

okonde=m	*ɟonom=le-n=a*
where=2SUBJ	born-ANT-ITR-IND

'Where were you born?'

(102) Mayurbhanj Ho

mandi	*ɟom-ke-ɖ=a=m*
rice	eat-PFV.TR-TR-IND=2.SUBJ

'You ate rice'

So the system of polypersonal verbs in Munda appears to be entirely independently derived from contemporary and historical South Asian contact history with Indo-Aryan and/or Dravidian languages, nor does it appear to be an archaic retention from an earlier Austroasiatic system, but does have parallels to systems found in other marginalized minority language groups that may reflect a more ancient areal norm also found in Kiranti and Burushaski. Whatever the explanation of the rise of such structures, it is quite clearly not due to contact with Aryan or Dravidian languages.

6.0 Inherited Features and Innovative Features in Munda Auxiliary Verb Constructions

Munda languages make extensive use of auxiliary verb constructions. Some of these appear to be quite old patterns with parallels in other Austroasiatic languages, some are of very recent origin and directly attributable to contact effects with Dravidian or Indo-Aryan languages, while there are still other formations that appear to be Munda-internal developments reflecting 'normal' processes from the perspective of the typology of inflection in auxiliary verb constructions (Anderson 2006, 2011).

One clearly old pattern is the use of a reduplicated lexical verb with a finite/inflected auxiliary (103). Of course, the syntactic order of Lexical Verb + Auxiliary Verb is an innovation in Munda; the reduplicated verb stem is likely old. This pattern is found in Plains Gtaʔ (104) and Remo (105), for example.

(103)Lexical verb$_{\text{<REDUPLICATED NON-FINITE FORM>}}$Auxiliary Verb$_{\text{<FULLY INFLECTED FINITE>}}$

(104)Plains Gtaʔ

ccoŋ n-ɖiŋ-e
RDPL~eat 1-IPFV=FUT
'I will eat'
(Mahapatra et al. 1989)

(107) a. Remo b. Remo(Field Notes)

bəba-ɖen-ti-ŋ
RDPL~slap-PROG-NPST-1
'I am slapping'

bəba-ɖen-ki=niŋ
RDPL~slap-PROG-PRF=1
'I was slapping'

Note that the above forms actually instantiate more than one historical process here. In Gtaʔ, the lexical verb is prosodically independent from the auxiliary, while in Remo the inflected auxiliary and the lexical verb have already been integrated into a single word ('univerbated'). While

synchronically a sequence of two TAM markers now in Remo, historically this was an auxiliary verb + a TAM marker (106).

(106) REDUPL-Verb-TAM_I-TAM_J(-SUBJ)<*REDUPL-Lexical Verb+Auxiliary Verb-TAM_J-(SUBJ)

Other Munda languages show univerbated AVCs that originally had a reduplicated lexical verb. Thus the element *lo*(*ʔ*), functionally a frequentative in Sora and a continuative in Kharia, arose from the fusing of an auxiliary verb and a reduplicated lexical verb (if monosyllabic). The same historical process underlies the development of the progressive in Juang as well.

(107) Sora (Ramamurti 1931: 28)

gu-gu-loː-te-n
RDPL-call-FREQ-NPST-ITR
'He calls frequently'

(108) Kharia(Malhotra 1982: 145)

gamgam-loʔ-ki-may
RDPL:talk-CONT-PST.I-PL
'They kept on talking'

(109) Juang (Pinnow 1960-ms: 122)

aiɲ ɟe'gɟe'g-nom-an
I RDPL.cry- PROG-PST.ITR
'I was weeping'

Hill Gtaʔ uses this pattern in the formation of its negative progressive (110). Probably something like this stands at the origin of the typologically unusual use of reduplication with no negative scope operator in the Aslian language Semaq Beri (111), where perhaps a now lost negative auxiliary was once used.

(110) Hill Gtaʔ

nœjŋ a-ná kəkéj n-a-ɖéj=tɛ
IOBJ-you RDPL~see 1-NEG-AUX=NPST
'I'm not looking at you'

(111) Semaq Beri (Kruspe 2015: 486)

kɛ gh-gəh kweh
3SG RDPL:IPFV~-not.share biscuit
'He isn't sharing his biscuits'

Another likely old formation in the history of complex predicates in Munda can be seen in the form of an unmarked lexical verb together with

an inflected auxiliary (112). Both Plains Gtaʔ (113) and Santali (114) reflect such a structure.

(112) Lexical Verb<Ø-MARKED NON-FINITE>Auxiliary Verb<FULLY INFLECTED FINITE>

(113) a. Plains Gtaʔ b. Plains Gtaʔ c. Plains Gtaʔ

coŋ n-læʔ-ge	*coŋ n-á-læʔ-ke*	*knweʔ-ɾæ gweʔ* ***we****-ge*
eat-1-PFV-PRF	eat 1-NEG-PFV-AOR	wife-3 REF dieAUX-PST.I
‘I have eaten’	‘I hadn’t eaten’	‘His wife had died’

(114) a. Santali (Bodding 1929: 277) b. Santali

ɟɔm baɽa-ke-t’-a-ko	*əgu hɔt’-ke-t’-ko-a-e*
eat AUX-AOR-TR-IND-PL	bring AUX-AOR-TR-PL-IND-3
‘They are done eating’	‘He brought them quickly’

In Santali, these are arguably univerbated synchronically [5], but they obviously remain two freestanding words in Plains Gtaʔ.
However, some clearly secondary developments due to language contact are attested when looking at auxiliary verb constructions in Munda languages. One of these is the use of a converb that is formally identical with a past tense marker as the construction-specific non-finite form of the lexical verb (115) in an auxiliary verb construction found, for example, in Remo (116) and Gutob (117).

(115) Lexical verb<PST.PRTCPL/CV NON-FINITE FORM>Auxiliary Verb<FULLY INFLECTED FINITE>

(116) Remo (Fernandez 1983) (117) Gutob (Hook 1991: 185)

baɖ-oʔ suŋ-oʔ -niŋ	*golgolte gai-gi ui-to*
slap-PST.II.CV AUX-PST.II-1	smoothly enter-PST.I.CVAUX-HAB
‘I finished slapping’	‘Smoothly it goes in..’

Another clearly secondary feature in an auxiliary verb construction can be seen in the use of the infinitive marker *-na* (arguably a loan from Indo-Aryan but possibly an internal development as well) in Kharia in the following AVC (118)-(119).

(118) Lexical verb$_{\text{<INFINTIVE NON-FINITE FORM>}}$Auxiliary Verb$_{\text{<FULLY INFLECTED FINITE>}}$

(119) i. Kharia (Malhotra 1982: 357)

kol-ob-ɲoʔ-ɖom-d^{h}ab-na	*laʔ-ki-kiyar*
REC-CAUS-eat-PASS-CEL-INF	IMPFV-PST:ITR-3DL

'They two were being fed by each other quickly'

It appears that Indo-Aryan influence in Kharia morphosyntax has increased over the past century if we compare data from the late 19th century to more recent data. In Banerjee (1893), the auxiliary *cuki* (120) appeared with an unmarked lexical verb (like Santali above) but this same auxiliary took an infinitive complement in Biligiri's data from the mid-20th century (121).

(120) Kharia

iŋ	*ɲoʔ*	*cuki-k-iŋ*
I	eat	COMPL-PST.I-1

'I have finished eating' (Banerjee 1893)

(121) Kharia

ɲoʔ-na	*cuki-k-iŋ*
eat-INF	COMPL-PST.I-1

'I have finished eating' (Biligiri 1965)

In Gorum (123), one finds a curious doubly inflected pattern (122) in certain auxiliary verb constructions where both tense and subject (and affectedness) appear with both the lexical verb and the auxiliary verb.

(122) Lexical verb$_{\text{<FULLY INFLECTED FORM>}}$ Auxiliary Verb$_{\text{<FULLY INFLECTED FINITE>}}$

(123) a. Gorum (Parengi)

miŋ	*ne-gaʔ-ru*	*ne-laʔ-ru*
I	1-eat-PST	1-AUX-PST

'I ate vigorously' (Aze 1973: 279)

b. Gorum (Parengi)

miŋ	*ne-adaʔ-ruʔ*	*ne-k-ruʔ*	
I	1-thirst-PST:AFF	1-AUX-PST:AFF	

'I was thirsty' (Aze 1973: 296)

Gutob may show similar formations as well, but note that the form in (124) could be explained differently as well, as there is promiscuous use of subject clitics found in Gutob, where multiple instantiations of the same agreement marker can be found across words in a clause, and this may be an example of this, not a doubled pattern, since the third element in the structure *ɖu-gu* does not bear subject inflection.

(124) Gutob (N. Zide 1997: 316)

niŋ	*ri-oʔ=niŋ=su*	*biɽ-oʔ=niŋ*	*beɽ-oʔ=niŋ*	*ɖu-gu*
I	bring-PST.II-1-SS	sow-PST.II-1	AUX-PST.II-1	AUX-PST.I

'[after] I brought (the seed) <and> I sowed it'

The closest analog to the Gorum situation is found in Dravidian languages also spoken in roughly the same region as Gorum where doubled subject inflection in AVCs is also attested, but reflecting more complicated split/doubled patterns sometimes, with subject doubled but tense or negation showing a split distribution appearing on either the lexical verb or the auxiliary verb alone, not both like subject. Such forms can be found for example in Muria Gondi (125) or Parji (126).

(125) Muria Gondi

*punno**n***	*atan*
know-NEG-**1**	AUX-PFV-**1**

'I didn't know' (Steever 1997: 290-1)

(126) Parji

*nil-t-**en***	*mẽ-d-**an***
stand-PST.PRTCPL-**1**	AUX-NPST-**1**

'I am standing, have stood up' (Steever 1988: 89)

Lastly, I mention here one final structure seen in auxiliary verb constructions in Munda languages and that is the split negative/subject pattern seen in Remo. The lexical verb takes the negative scope operator, and the auxiliary rather takes the subject marker (127). Such a pattern in Remo (130) is also found in Kiranti languages of Nepal, e.g., Thulung (128) or Camling (129). But as this is among the most common inflectional splits seen cross-linguistically (Anderson 2006), there is no reason to consider this anything other than a normal internal development within Remo.

(127) Split pattern:NEGATIVE-Lexical Verb Auxiliary Verb-SUBJECT

(128) Thulung

mi-pe-thiŋa *bu-ŋa*
NEG-eat-CV AUX-1
'I have not eaten' (Ebert 2003a: 513)

(129) Camling

mi-tim *ŋas-i-e*
NEG-meet AUX-1PL-NPST
'we have not met'
(Ebert 2003b: 541)

(130) Remo

a-sap *ɖen-gi-t-iŋ*
NEG-come PROG-PST.I-NPST-1
'I have not been coming'
(Fernandez 1983)

7.0 Summary

Overall, it can thus be said that some features of the modern Munda languages such as verb-final and GEN N order with nominal possessors marked by a genitive case clearly reflect a secondary development in Munda that aligns these languages with South Asian areal norms syntactically and likely occurred already at the proto-Munda stage. But many other features show either later contact effects or rather instead reflect inheritances from an earlier period in the Munda languages that aligns these with their eastern linguistic cousins in the Austroasiatic phylum. Also, some features of Munda indeed find neither analogs in the languages in South Asia they are presently in contact with nor in other Austroasiatic languages, and thus should be considered Munda-internal changes not triggered by contact, or if contact-driven, reflect an earlier period of contact in the languages that pre-dates that of the Arayo-Dravidian period in South Asia, which the Munda speakers may pre-date the arrival of in the Subcontinent. In short, the overly facile explanation of South Asian contact effects in the history of Munda languages allegedly triggered by a shift in the fundamental 'rhythmic holism' of the proto-Munda language from iambic to trochaic rhythm that completely 'reset', as it were, the parameter settings of the language away from its putative ancestral type (ostensibly something akin to the present-day Mainland Southeast Asia type) towards the present-day South Asian areal type must be replaced by a significantly more nuanced historical approach. This approach accepts i) that South Asian features have been accrued at different points in time and thus by different Munda languages individually, and differently in some occasions, and at intermediate-level stages or at the proto-Munda level as well or indeed very recently, ii) that

Munda languages can and do offer insights into the earlier history of Austroasiatic languages by having likely better preserved some older features later erased in most of the languages remaining in Mainland Southeast Asia (noun incorporation, possibly the objective case prefix), and iii) that there are features of Munda for which neither analogs in other branches of the Austroasiatic languages nor in other major South Asian genetic units will be found nor should be sought, or if they are, they should be sought elsewhere to an earlier, now peripherally located and marginalized typological profile, that may pre-date the Aryan and Dravidian migrations into South Asia (polypersonal verbs).

As we shed previous constraints on the analysis of the Munda languages (and other Austroasiatic languages too), both instrumental and meta-theoretical, we can move to a better understanding of what Munda languages are really like and how they really became the way they are, and in turn, gain both a more refined understanding of what Proto-Austroasiatic languages may have been like, and what the different contact dynamics at play were at different pre-historical and historical periods, and what language groups may have been involved in the accrual of contact-derived features in both South Asia and Southeast Asian branches of the Austroasiatic stock.

Abbreviations

ABL	Ablative	IRR	Irrealis
ACC	Accusative	ITR	Intransitive
ACT	Active	LOC	Locative
ADJVZR	Adjectivalizer	M	Masculine
ADS	Adessive	MDL	Middle
ALL	Allative	NEG	Negative
ANT	Anterior	NPST	Non-Past
AOR	Aorist	NSFX	Noun Suffix
APPL	Applicative	OBJ	Object[ive]
ASP	Aspect	OBL	Oblique
AUX	Auxiliary	PFV	Perfective
BEN	Benefactive	PHB	Prohibitive
CAUS	Causative	PL	Plural
CLSSFR	Classifier	PLURACT	Pluractional
COND	Conditional	PRG	Progressive
COP	Copula	PRON	Pronoun
CV	Converb	PRS	Present
DAT	Dative	PST	Past
DECL	Declarative	PST.I	Past Series-i
DEF	Definite	PST.II	Past Series-ii

DESID	Desiderative	PSV	Passive
DIR	Directional	PURP	Purposive
DL	Dual	QUOT	Quotative
DS	Different Subject	RDPL	Reduplication
EMPH	Emphatic	RECIP	Reciprocal
EVID	Evidential	REF	Referential
EXCL	Exclusive	RFLXV	Reflexive
FEM	Feminine	RLS	Realis
FIN	Finite	RLS	Realis
FOC	Focus	SG	Singular
FUT	Future	SUBJ	Subject
GEN	Genitive	TAM	Tense-Mood-Aspect
HAB	Habitual	TR	Transitive
HUM	Human	1	1st person
IMP	Imperative	2	2nd person
INAN	Inanimate	3	3rd person
IPFV	Imperfective		

Notes

[1] Note also that there is variation between the presence and absence of the minor/pre-syllable in certain words in Thavung-So [**Vietic**], Thailand], e.g. *[pha]lə̂ŋ* 'to forget', *[ʔa]pên* 'fly' *[ka]loːŋ* 'coffin' (Premsrirat 1996: 168).

[2] Note that Oḍia shows similar preverbal + postverbal patterns, but the latter is common in many finite verbs, the former more restricted (Masica 1993: 391-392).

(a) Odia

mũ	*jāe=ni*	*tume*	*jibo=ni*	*āpɔnɔ*	*gɔle=ni*
1SG	go:PRS=NEG	2SG	go:FUT=NEG	2.HON	go:PST=NEG
'I don't go'		'You will not go'		'You (hon) did not go'	

(b) Odia

nɔ-thī-li	*nɔ-thi-bi*
NEG-be-PST.1	NEG-be-FUT.1
'I was not'	'I won't' (Pattanayak and Das 1972: 131)

[3] For example, the typologically bizarre pattern seen in Sora where transitive agents are incorporated in the pattern VA, reflecting the proposed syntactic order of Proto-Austroasiatic.

(c) Sora

*ɲem=**bud**-t-am*	*ɲem=**bun**-t-iɲ*
seize-**bear**-NPST-UND	seize-**pig**-NPST-1UND
'the bear will seize you'	'the pig will seize me' (Anderson 2017: 946)

[4] This functional conflation is typical of Central Dravidian as a whole too. Also, some formal similarity in markers exists: Kuvi (Reddy 1979) *–ki* (also Telugu, Masica 2007), Kui *–gi/-ki* (Krishnamurti 2003: 220, 232). How or whether these should be considered in the discussion here remain an open question and a topic for future research.

[5] Also a fusing of auxiliary verbs with unmarked lexical verbs is what underlies the development of the Kherwarian perfect series of inflections (Anderson 2007a), as seen in the following forms from Ho (repeating 101-102 above):

(d) Mayurbhanj Ho
okonde=m *ɟonom-**le-n**-a*
where=2 born-ANT-ITR/MDL-IND
'Where were you born?' (Field Notes)

(e) Mayurbhanj Ho
mandi *ɟom-**ke-ɖ**-a=m*
rice eat-PFV.TR-TR/ACT-IND=2
'You ate rice' (Field Notes)

References

Alves, Mark J. "Ruc and the other minor Vietic languages: Linguistic strands between Vietnamese and the rest of the Mon-Khmer family." In *SEALS VII Papers from the 7th Annual Meeting of the Southeast Asian Linguistics Society (SEALS VII),* edited by Karen L. Adams, Thomas John Hudak, and F.K. Lehman, 3-19. Arizona: Arizona State University, 2003.

—. *A Pacoh grammar*. Canberra: Pacific Linguistics, 2004.

—. "The Vieto-Katuic hypothesis." In *SEALS XV Papers from the 15th Annual Meeting of the Southeast Asian Linguistics Society (SEALS-XV)*, edited by P. Sidwell, 169-176.Canberra: Pacific Linguistics, 2005.

—. "Pacoh Pronouns and Grammaticalization clines." In *SEALS XIII Papers from the 13th Annual Meeting of the Southeast Asian Linguistics Society*, edited by I. Shoichi et al., 1-12. Canberra: Pacific Linguistics, 2007.

—. "Grammatical Functions in Mon-Khmer Morphology." Paper presented at the 23rdAnnual Meeting of the Southeast Asian Linguistics Society, Bangkok, Thailand, 2013.

—. "Mon-Khmer." In *The Oxford Handbook of Derivational Morphology* edited by R. Leiber, and Pavol Stekauer, 520-544. Oxford: Oxford University Press, 2014.

—a. "Morphological functions among Mon-Khmer languages." In *Languages of Mainland Southeast Asia: The State of the Art*, edited by N. Enfield and B. Comrie, 531-557, Berlin: de Gruyter, 2015.

—b. "Pacoh." In *The Handbook of Austroasiatic Languages*, edited by Mathias Jenny, and Paul Sidwell, 881-906.Leiden: Brill Publishers, 2015.

Anderson, Gregory D. S. *Xakas*. München: Lincom, 1998.
—. "A New Classification of Munda: Evidence from Comparative Verb Morphology." *Indian Linguistics* 62 (2001): 27-42.
—. "Advances in Proto-Munda Reconstruction." *Mon-Khmer Studies* 34(2004): 159-184.
—. *Auxiliary Verb Constructions*. Oxford: Oxford University Press, 2006.
—a. *The Munda Verb. Typological Perspectives.* Berlin: Mouton de Gruyter (Trends in Linguistics, Studies, and Monographs, 174), 2007.
—b. "Burushaski Morphology." In *Morphologies of Asia and Africa*, edited by Alan Kaye, 1233-1275.Winona Lake: Eisenbrauns, 2007.
—. "Gta?." In *The Munda Languages*, edited by Gregory D.S. Anderson, 682-763. London & New York: Routledge, 2008.
—, ed. *The Munda Languages*. London & New York: Routledge, 2008.
—a. "Auxiliary Verb Constructions (and Other Complex Predicate Types): A Functional-Constructional Typology." *Language and Linguistics Compass* 5, no.11 (2011): 795-828.
—b. "Auxiliary Verb Constructions in the Languages of Africa." *Studies in African Linguistics* 40, no. 1-2(2011):1-409.
—. "What Munda languages are really like." Keynote address presented at the All India Conference of Linguists (AICL-34), Shillong, India November 1-3,2012. Keynote address.
—a. "Overview of the Munda languages." In *The Handbook of Austroasiatic Languages*, *Volume 1*, edited by Mathias Jenny and Paul Sidwell, 364-414. Amsterdam: Brill, 2015.
—b. "Prosody, phonological domains and the structure of roots, stems and words in the Munda languages in a comparative/historical light." *Journal of South Asian Languages and Linguistics* 2, no.2 (2015): 163-183.
—. "Do Koraput Munda, Lower Munda, and even South Munda really exist? Once more on the still unresolved classification of the Munda languages." In *Multilingualism and Multiculturalism. Perceptions, Practices and Policy*, edited by S. Pattanayak, C. Pattanayak, and J. M. Bayer, 313-334. Delhi: Orient Blackswan, 2016.
—. "Polysynthesis in Sora, with special reference to noun incorporation." In *Handbook of Polysynthesis*, edited by Michael Fortescue, Marianne Mithun, and Nicholas Evans,930-947. Oxford: Oxford University Press, 2017.
—. "Proto-Munda Prosody, Morphotactics and Morphosyntax in South Asian and Austroasiatic Contexts." In *Austroasiatic Syntax in Areal and Diachronic Perspectives*, edited by Mathias Jenny, Paul Sidwell, and Mark Alves, 157-197. Leiden: Brill,2020.

—. In preparation. *The Gta? Language. Hill Gta? and Plains Gta? Grammar, Texts, Lexicon.*

Anderson, Gregory D. S., and Opino Gomango. "On the current status and state of Juray in the Sora-Juray cluster." In *Proceedings of the Foundation of Endangered Languages* (FEL-XX) edited by Nicholas P. Ostler and Panchanan Mohanty,103-109. Bath: FEL, 2017.

Anderson, Gregory D. S., Opino Gomango and Luke Horo. In preparation. *A comparative survey of Sora dialects.*

Anderson, Gregory D. S and K. David Harrison. *Tyvan.* München: Lincom, 1999.

Anderson, Gregory D. S., and K. David Harrison. "Sora." In *The Munda Languages*, edited by Gregory D.S. Anderson, 299-380.London &New York: Routledge, 2008.

Anderson, Gregory D. S., and Bikram Jora. "Negation, TAM and person-indexing interdependencies in the Munda languages: a preliminary report." In *Papers from the Seventh International Conference on Austroasiatic Linguistics*, edited by Hiram Ring and Felix Rau, 36-59, Hawai: University of Hawai'i, 2018.

Arsenault, Paul. *"Retroflex consonant harmony in South Asia."* Ph.D. Dissertation. University of Toronto, 2012.

—. "Retroflexion in South Asia: Typological, Genetic, and Areal Patterns." *Journal of South Asian Languages and Linguistics*4, no.1 (2017):1-53.

Aze, F. R. "Clause Patterns in Parengi-Gorum." In *Patterns in clause, sentence, discourse in selected languages of India and Nepal* 1, edited by Robert. L. Trail, 235-312. Norman, OK: SIL, 1973.

Bätscher, Kevin. 2015. "Mlabri." In *The Handbook of Austroasiatic Languages*, *Volume II,* edited by Mathias, Jenny and Paul Sidwell, 1003-1030. Amsterdam: Brill, 2015

Banerjee, G. Ch. *Introduction to the Kharia Language*. Calcutta: Bengal Secretariat Press, 1894.

Banker, Elizabeth M. a. "Bahnar affixation." *Mon-Khmer Studies* 1(1964): 99-117.

—b. "Bahnar reduplication." *Mon-Khmer Studies* 1(1964):119-134.

Bauer, Christian. "Morphology and syntax of spoken Mon." Ph.D. Dissertation. University of London, 1982.

—. "Recovering extracted infixes in Middle Khmer." *Mon-Khmer Studies* 15(1986): 155-164.

—. "The Verb in Spoken Mon." *Mon-Khmer Studies* 15(1989): 87-110.

— "Reanalyzing reanalyses in Katuic and Bahnaric." *Mon-Khmer Studies* 16-17(1990): 143-182.

Benjamin, Geoffrey. "An outline of Temiar Grammar." In *Austroasiatic Studies Part I*, edited by P. N. Jenner, L. C. Thompson, and S. Starosta, 129-188, (Oceanic Linguistics Special Publication No. 13). Honolulu: University of Hawaii Press, 1976.

—. "Temiar Morphology (and Related Features): A View from the Field." Paper presented at the Fifth International Conference Austroasiatic Linguistics 5. Mahidol University, Salaya, Thailand, 2011

Berger, Hermann. *Die Burushaski-Sprache von Hunza und Nager*. 3 vols. Wiesbaden: Harrassowitz Verlag, 1998.

Biligiri, H. S. *Kharia: phonology, grammar, vocabulary*. Poona: Deccan College, 1965.

Bisang, Walter. "Modern Khmer." In *The Handbook of Austroasiatic Languages*. *Volume II,* edited by Mathias Jenny and Paul Sidwell, 677-716. Amsterdam: Brill, 2015

Bishop, N. "A preliminary description of Kensiw (Maniq) Phonology."*Mon-Khmer Studies* 25 (1996): 227-254.

Bishop, N., and M. M. Peterson. "Kensiw glossary." *Mon-Khmer Studies* 23 (1994): 163-196.

Bodding, P. O. *Santal Grammar for Beginners*. Dumka: Santal Mission of Northern Churches, 1929.

Bos, Kees Jan, and Paul Sidwell. "Kui Ntua." In *The Handbook of Austroasiatic Languages*. *Volume II,* edited by Mathias Jenny and Paul Sidwell, 837-880. Amsterdam: Brill, 2015

Brunelle, Marc. "Vietnamese." In *The Handbook of Austroasiatic Languages*. *Volume II,* edited by Mathias Jenny and Paul Sidwell, 909-953. Amsterdam: Brill, 2015

Brunelle, Marc, and Pittayawat Pittyaporn. "Phonologically-constrained change. The role of the foot in monosyllabization and rhythmic shifts in Mainland Southeast Asia." *Diachronica* 29, no.4 (2012): 411-433.

Burenhult, Niclas. "A Grammar of Jahai." Ph.D. Dissertation. Dept. of Linguistics and Phonetics, Lund University, 2002.

Butler, Becky. "Bunong." In *The Handbook of Austroasiatic Languages*. *Volume II,* edited by Mathias Jenny and Paul Sidwell, 719-745. Amsterdam: Brill, 2015

Costello, Nancy A. "Affixes in Katu." *Mon-Khmer Studies* 2(1966): 63-86.

—. "Affixes in Katu of the Lao P.D.R." *Mon-Khmer Studies* 28(1998): 31-42.

—. "Aspect and Tense in Katu of the Lao P.D.R." *Mon-Khmer Studies* 31(2001): 121-125.

Daladier, A. "Pnaric-War-Lyngngam and Khasi as a branch of Pnaric."

Journal of Southeast Asian Linguistics 4 (2011): 169-206.
Das, A. R. *A Study on the Nicobarese Language*. Calcutta: Anthropological Survey of India, 1977.
Dasgupta, Kamalesh. *A phrase book in Miju*. Shillong: Government of Arunachal Pradesh, 1977.
Deepadung, Sujartilak, Ampika Rattanapitak and Supakit Buakaw. "Dara'ang Palaung." In *The Handbook of Austroasiatic Languages*. *Volume II,* edited by Mathias Jenny and Paul Sidwell, 1065-1103. Amsterdam: Brill, 2015.
Diffloth, Gérard. "Jah-Hut: An Austroasiatic language of Malaysia." In *South-East Asian Linguistic Studies*, Vol. 2, edited by Nguyen Dang Liem, 73–118. Canberra: Pacific Linguistics, 1976.
—. "Proto-Austroasiatic Creaky Voice." *Mon-Khmer Studies* 15 (1989): 139-54.
—. "Palaungic vowels in Mon-Khmer perspective." In *Austroasiatic Languages, Essays in honour of HL Shorto* edited by J.H.C.S. Davidson,13-38.London: School of Oriental and African Studies,1991.
Donegan, Patricia J. "Rhythm and vocalic drift in Munda and Mon-Khmer." *Linguistics in the Tibeto-Burman Area* 16, no.1 (1993): 1-43.
Donegan, Patricia J. & David Stampe. "Rhythm and the holistic organization of language structure." In *The Interplay of Phonology, Morphology, and Syntax*, edited by John Richardson, M. Marks, and A. Chukerman, 337–353. Chicago: Chicago Linguistic Society, 1983.
Donegan, Patricia J. & David Stampe. 2004. "Rhythm and synthetic drift of Munda." In *Yearbook of South Asian Languages and Linguistics* edited by R. Singh, 3-36. Delhi: Sage Publication.
van Driem, G. *A Grammar of Limbu*. Berlin: Mouton de Gruyter, 1987.
Dryer, Matthew S. "Word order in Sino-Tibetan languages from a typological and geographical perspective." In *Sino-Tibetan Languages*, edited by Graham Thurgood and Randy LaPolla, 43-55 Richmond: Curzon Press, 2003.
—. "Word order in Tibeto-Burman languages." *Linguistics of the Tibeto-Burman Area* 31, no.1 (2008): 1-83.
Ebert, K. a. "Kiranti languages: an overview." In *Sino-Tibetan Languages*, edited by Graham Thurgood and Randy LaPolla, 505-517. Richmond: Curzon Press, 2003.
—b. "Camling." In *Sino-Tibetan Languages*, edited by Graham Thurgood and Randy LaPolla, 533-545. Richmond: Curzon Press, 2003
Edmondson, J. "English-Bolyu Glossary." *Mon-Khmer Studies* 24 (1995): 133-160.

Enfield Nick J. and Gérard Diffloth. “Phonology and sketch grammar of Kri, a Vietic language of Laos.” *Cahiers de Linguistique - Asie Orientale* (CLAO) 38, no.1(2009): 3-69.

Fernandez, F. “A Grammatical Sketch of Remo.” Ph.D. dissertation University of North Carolina,1968.

— “The morphology of the Remo (Bonda) verb.” *International Journal of Dravidian Linguistics* 12, no.1 (1983): 15–45.

François, Alexandre. “Trees, Waves, and Linkages. Models of Language Diversification.” In *The Routledge Handbook of Historical Linguistics,* edited by Claire Bowern and Bethwyn Evans, 161-189. London: Routledge, 2014.

Gradin D. “Word affixation in Jeh.” *Mon-Khmer Studies* 5(1976): 25-42.

Hook, P. E. “The Compound Verb in Munda: An Areal Overview”. *Language Sciences* 13, no.2 (1991): 181-95.

Horo, Luke and Priyankoo Sarmah. “Acoustic analysis of vowels in Assam Sora.” In *Northeast Indian Linguistics 7,* edited by L. Konnerth et al., 69-86. Canberra: ANU, 2015.

—. “A Phonetic Description of Assam Sora.” Ph.D. Dissertation, Indian Institute of Technology, Guwahati, 2017.

Jenner, P. and S. Pou “A lexicon of Khmer morphology.” *Mon-Khmer Studies* 9-10(1982): 1-517.

Jenny, Mathias. “New infixes in spoken Mon.” *Mon Khmer Studies* 33(2003): 183-194.

—. *The verb system of Mon*. Zurich: ASAS, 2005.

—, and Patrick McCormick. “Old Mon.” In *The Handbook of Austroasiatic Languages*. *Volume II, edited* by Mathias Jenny and Paul Sidwell, 519-552. Amsterdam: Brill, 2015**.**

— “Modern Mon.” In *The Handbook of Austroasiatic Languages*. *Volume II,* edited by Mathias Jenny and Paul Sidwell, 553-600. Amsterdam: Brill, 2015

—, Tobias Weber, and Rachel Weymuth. “The Austroasiatic Languages: A Typological Overview.” In *The Handbook of Austroasiatic Languages*. *Volume I*, edited by Mathias Jenny and Paul Sidwell, 13-143. Amsterdam: Brill, 2015

— et al. b. “Syntactic typology of Austroasiatic.” Paper presented at ICAAL, Siem Reap, 2015.

—, and Paul Sidwell, eds. *The Handbook of Austroasiatic Languages*. 2 volumes. Leiden: Brill, 2015.

—, “Verb Initial Structures in Austroasiatic Languages.” In *Austroasiatic Syntax in Areal and Diachronic Perspectives*, edited by Mathias Jenny, Paul Sidwell, and Mark Alves,21-45. Leiden: Brill,2020.

Jordan-Horstmann, Monika. *Sadani. A Bhojpuri Dialect Spoken in Chotanagpur*. Wiesbaden: Otto Harrassowitz, 1969.

Kiran, Savita and John Peterson. "Sadani/Sadri." In *Новые индоарийские языки***.** Ред. кол. Л.И. Куликов, Т.И.Оранская, А.Ю. Русаков. - "Языки мира", т. 16, Российская Академия Наук, Институт Языкознания. Москва: Academia. 367-379, 2011.

Krishnamurti, B., *The Dravidian Languages*. Cambridge: Cambridge University Press, 2003.

Kruspe, Nicole. *A Grammar of Semelai*. Cambridge University Press, 2004.

—. "Semaq Beri." In *The Handbook of Austroasiatic Languages*. *Volume II,* edited by Mathias Jenny and Paul Sidwell, 475-516. Amsterdam: Brill, 2015

Kruspe, Nicole, Niclas Burenhult and Ewelina Wnuk. "Northern Aslian." In *The Handbook of Austroasiatic Languages*. *Volume I*, edited by Mathias Jenny and Paul Sidwell, 419-474. Amsterdam: Brill, 2015

Li, Jinfang. "Bugan: A New Mon-Khmer Language of Yunnan Province, China." *Mon-Khmer Studies* 26(1996): 135-159.

Li, Jinfang, and Yongxian Luo. "Bugan." In *The Handbook of Austroasiatic Languages*. *Volume II,* edited by Mathias Jenny and Paul Sidwell, 1033-1062. Amsterdam: Brill, 2015

Mahapatra, K., Dobek Pujari and P. K. Panda. *Didayi* (In Oriya). Academy of Tribal Dialects and Culture, Government of Orissa, Bhubaneshwar, 1989

Mahapatra, K and N. H. Zide. n.d. *Gtaʔ texts*. Unpublished manuscript.

Malhotra, V. "The structure of Kharia: a study in linguistic typology and change." Ph.D. dissertation. Jawaharlal Nehru University, 1982.

Masica, Colin P. *The Indo-Aryan Languages*. Cambridge: Cambridge University Press, 1993.

—. "Postverbal Subjects in Telugu and Other Languages." In *Old and New Perspectives on South Asian Languages. Grammar and Semantics,* edited by Colin P. Masica, 246-273. Delhi: Motilal Banarsidass Publishers, 2007.

Matisoff, James. "Sino-Tibetan linguistics: Present state and future prospects." *Annual Review of Anthropology* 20(1991): 469-504.

—a. "Aslian: Mon-Khmer of the Malay Peninsula." *Mon-Khmer Studies* 33(2003): 1-58.

—b. *Handbook of Proto-Tibeto-Burman.* University of California Publications in Linguistics 135. Berkeley: University of California Press, 2003.

Milne, Leslie. *An elementary Palaung grammar*. London: Oxford University Press, 1921.

Mohan, Shailendra. “Noun morphology in Nihali and Korku, a comparative account.” *Bulletin of the Deccan College Research Institute* Vol. 75 (2015):185-194.

Nagaraja, K. S. “Khasi dialects: a typological consideration.” *Mon-Khmer Studies* 23(1993): 1-10.

—. *Korku Language. Grammar, Texts, Vocabulary*. Tokyo: Tokyo University of Foreign Studies, Institute for the Study of Languages and Cultures of Asia and Africa, 1999.

—a. “Standard Khasi.” In *The Handbook of Austroasiatic Languages*. *Volume II,* edited by Mathias Jenny and Paul Sidwell,1145-1185. Amsterdam: Brill, 2015

—b. *Nihali Language*. Mysore: Central Institute of Indian Languages, 2015.

Nai Pan Hla. “A Comparative Study of Old Mon Epigraphy and Modern Mon.” *Oceanic Linguistics Special Publications* No. 13, Austroasiatic Studies Part II (1976): 891-918.

Nowrangi, P. S. *A simple Sadāni grammar*. Ranchi: D. S. S. Book Depot, 1956.

Olsen, Neil H. “Koho-Sre.” In *The Handbook of Austroasiatic Languages*. *Volume II,* edited by Mathias Jenny and Paul Sidwell,746-788. Amsterdam: Brill, 2015

Omar, Asmah Haji. “The verb in Kentakbong.” In *Austroasiatic Studies Part II*, edited by P. N. Jenner, L. C. Thompson, and S. Starosta, 951-960, (Oceanic Linguistics Special Publication No. 13). Honolulu: University of Hawaii Press, 1975.

Osada, Toshiki. “Mundari.” In *The Munda Languages*, edited by Gregory D.S.Anderson, 99-164.London &New York: Routledge, 2008.

Patnaik, M. “Juang.” In *The Munda Languages*, edited by Gregory D.S. Anderson. 508-556. London &New York: Routledge, 2008.

Pattanayak, D. P. and G. N. Das. *Conversational Oriya*. Mysore: Smt Sulakshana Pattanayak, 1972.

Peiros, I. 1996. *Katuic Comparative Dictionary*. Pacific Linguistics C-132. Canberra: Australian National University, 1996.

Peterson, John M. “Kharia.” In *The Munda Languages*, edited by Gregory D.S. Anderson, 434-507.London & New York: Routledge, 2008

—. *Grammar of Kharia*. Leiden: Brill, 2011.

Pinnow, Heinz-Jürgen. *Versuch einer historischen Lautlehre der Kharia-Sprache.*Wiesbaden: Otto Harrassowitz, 1959.

—. -ms. *Beiträge zur Kenntnis der Juang-Sprache*. Unpublished-ms, Berlin, 1960.

—. "Über den Ursprung der voneinanderabweichenden Struktur der Munda und Khmer-Nikobar Sprachen." *Indo-Iranian Journal* 4 (1960):81-103

—. "The position of the Munda languages within the Austroasiatic language family." In *Linguistic Comparison in Southeast Asia and the Pacific,* edited byH.L. Shorto, 140-152.London: SOAS, 1963.

—. "A Comparative Study of the Verb in the Munda Languages." In *Studies in Comparative Austroasiatic Linguistics,* edited by Norman H. Zide, 96-193. The Hague Mouton (Indo-Iranian Monographs, V), 1966.

Post, Mark W. "Prosody and typological drift in Austroasiatic and Tibeto-Burman: Against "Sinosphere" and "Indosphere." In *Austroasiatic Studies*, edited by S. Srichampa, Paul Sidwell, and Kenneth Gregerson,198-221, Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics,2011.

Premsrirat, Suwilai. "Phonological characteristics of So (Thavung), a Vietic language of Thailand." *Mon-Khmer Studies* 26 (1996): 161-176,

Premsrirat, Suwilai, and Nattamon Rojankul. "Chong." In *The Handbook of Austroasiatic Languages*. *Volume II,* edited by Mathias Jenny and Paul Sidwell,603-640. Amsterdam: Brill, 2015

Pucilowski, Anna. "Complementation in Ho (North Munda)." *Mon-Khmer Studies* 41(2012): 20-31.

Pulu, Jimi. *A handbook on Idu Mishmi language*. Itanagar: Department of Cultural Affairs, Directorate of Research, Government of Arunachal Pradesh, 2002.

Radhakrishnan, R. *The Nancowry word: phonology, affixal morphology and roots of a Nicobarese language.* Edmonton: Linguistic Research, Inc, 1981.

Ramamurti, G. V.*A Manual of the So:ra: (or Savara) Language*. Madras: Govt. Press, 1931.

Reddy, Joy. *Kuvi Grammar*. CIIL Grammar Series 4. Mysore: Central Institute of Indian Languages, 1979.

Ring, Hiram. "Pnar." In *The Handbook of Austroasiatic Languages*. *Volume II,* edited by Mathias Jenny and Paul Sidwell, 1186-1226. Amsterdam: Brill, 2015.

Ring, Hiram, and Gregory D. S. Anderson. "On prosodic structures in Austroasiatic diachrony: 'Rhythmic holism' revisited in light of preliminary acoustic studies on Khasian and Munda." In *Papers from the Seventh International Conference on Austroasiatic Linguistics*, edited by Hiram Ring and Felix Rau, 1-35, Hawai: University of Hawai',2018.

Ross, Malcolm D. *Proto-Oceanic and the Austronesian languages of Western Melanesia*. Canberra: Australian National University, 1988.

—. "Social networks and kinds of speech-community event." In *Archaeology and language. Volume 1: Theoretical and methodological orientations* edited by Roger Blench and Matthew Spriggs, 209-261. London: Routledge, 1997.

—. "Calquing and metatypy." *Journal of Language Contact, Thema* 1(2007): 116-43.

Schiller E. "Khmer nominalizing and causativizing infixes." In *Papers from the Second Annual Meeting of the Southeast Asian Linguistics Society,* edited by K.L. Adams and T.J. Hudak, 309-326. Arizona State University, Program for Southeast Asian Studies, 1994.

Shorto, H. "The Vocalism of Proto-Mon-Khmer." *Oceanic Linguistics Special Publication*13, no.2 (1976): 1041-1067.

—. *A Mon-Khmer comparative dictionary*. Canberra: Australian National University, 2006.

Si, Aung. "Danau." In *The Handbook of Austroasiatic Languages*. *Volume II,* edited by Mathias Jenny and Paul Sidwell, 1104-1141. Amsterdam: Brill, 2015

Sidwell, Paul. "Issues in the morphological reconstruction of Proto-Mon-Khmer." In *Morphology and Language History: In honour of Harold Koch*, edited by Claire Bowern, Bethwyn Evans and Luisa Miceli, 251–265. Canberra, 2008.

—. *Classifying the Austroasiatic Languages: History and State of the Art*. München: Lincom, 2009.

—. "Proto-Khasian and Khasi-Palaungic." *Journal of Southeast Asian Linguistics* 4(2011): 144-168.

—. ms. Diversity, discontinuity, and asymmetry in the typological restructuring of Mainland Southeast Asian languages.

—. "Issues in Austroasiatic Classification." *Languages and Linguistics Compass* 7/8(2013): 437-457.

—a. "Austroasiatic Classification." In *The Handbook of Austroasiatic Languages*. *Volume I*, edited by Mathias Jenny and Paul Sidwell,114-220. Amsterdam: Brill, 2015.

—b. "Car Nicobarese." In *The Handbook of Austroasiatic Languages*. *Volume II*, edited by Mathias Jenny and Paul Sidwell,1229-1265. Amsterdam: Brill, 2015.

—c. "Local drift and areal convergence in the restructuring of MSEA." In *Languages in Mainland South East Asia: The state of the art*, edited by N. Enfield and B. Comrie, 51-81. Berlin: de Gruyter, 2015.

Sidwell, Paul, and Felix Rau. "Austroasiatic Comparative-Historical Reconstruction: An Overview." In *The Handbook of Austroasiatic Languages*. *Volume I*, edited by Mathias Jenny and Paul Sidwell, 221-363. Amsterdam: Brill, 2015.

Smith, Kenneth D. "Sedang Affixation." *Mon-Khmer Studies* 3(1969): 108-124.

Smith, R.L. "Reduplication in Ngeq." *Mon-Khmer Studies* 4(1973): 85-111.

Solntseva, Nina. "Case-marked Pronouns in the Taoih Language." *Mon-Khmer Studies* 26(1996): 33-36.

Steever, S. B. "Gondi." In *The Dravidian Languages*, edited by S. B. Steever,270-297. London: Routledge, 1997.

—. *The Serial Verb Construction in the Dravidian Languages*. Delhi: Motilal Banarsidas, 1988.

Svantesson, Jan-Olof. *Kammu phonology and morphology*. Sweden: CWK Gleerup, 1983.

Tayeng, A. *Milang Phrase Book*. Shillong: Directorate of Information, Arunachal Pradesh, 1976.

Thomas, D. M., "Chrau affixes." *Mon-Khmer Studies* 3(1969): 90-107.

—. "The Instrument/Locative and Goal Affix -N- in Surin Khmer." *Mon-Khmer Studies 16*-17(1990): 85-98.

Thurgood, Graham and Randy La Polla, eds. *Sino-Tibetan Languages*. Routledge Language Family Series. London/New York: Routledge, 2003.

Watson, Richard. "A Case for Clitics in Pacoh." In *Austroasiatic Studies: Papers from ICAAL4. Mon-Khmer Studies Special Issue No. 3*. edited by Sophana Srichampa, Paul Sidwell, and Kenneth Gregerson, 222-232. Dallas, SIL International; Salaya, Mahidol University; Canberra, Pacific Linguistics, 2011

Watson, Saundra. "Personal pronouns in Pacoh." *Mon-Khmer Studies* 1(1964):81-97.

—. "Verbal affixation in Pacoh." *Mon-Khmer Studies* 2(1966):15-30.

Watters, D., "Notes on Kusunda grammar: A language isolate of Nepal." *Himalayan Linguistics Archive* 3 (2006): 1-182.

You Sey. "Some Old Khmer affixation." *Mon-Khmer Studies* 5(1976): 85-95.

Zide, Norman H., "Gutob pronominal clitics and related phenomenon elsewhere in Gutob-Remo-Gtaʔ." In *Languages of tribal and indigenous people of India: The ethnic space*, edited by Anvita Abbi, 307-334. New Delhi: MLBD, 1997.

CHAPTER TWO

INFIXATION IN MUNDA AND ITS AUSTROASIATIC LEGACY

ARUN GHOSH

1.0 Introduction

Infixation is a linguistic process that may operate in both morphological and syntactic domains. In the morphological domain, it engages in derivation while in the syntactic domain it engages in grammatical categorization. In the syntactic domain, grammatical elements may be infixed in the verb-words to express tense/aspect/mood, person, or number, as in

(1) Sa. *dal=ko=a=ɲ* 'I shall beat them.'
beat=3pl.O=F=1sg.S

(2) Mu. *ja merom nam=a=i=me* 'Look for any goat for him/her'
any goat get=A=3sg.O=2sg (Osada 1992:79)

In example 1 in Santali 3pl.object and in example 2 in Mundari 3sg.object and the applicative are infixed. In the morphological domain, the process of infixation engages in word derivation from an existing base (root or stem) without any syntactic determination. Both types are treated within infixation in some literature, although there is a clear cut dichotomy between the two. Here in this chapter, we have considered only those infixes which operate in the morphological domain as class-changing devices and/or sub-class producing devices. In that sense, our focus is on infixation as a derivational process. We follow Blevins' consideration of Stump (2005):

> "If infixation imposes part-of-speech membership, then we treat it as derivational. If an operation is complete and semantically regular, it is

> usually inflectional, not derivational. If it is syntactically determined, it is also inflectional, not derivational." (Blevins 2012: 7)

The cross-linguistically derivational operation normally precedes inflectional operation as the derived forms take the inflections after the derivational operation is over, as in

(3) Sa. *ɛ<tɔ>hɔp=re*[1] *galmarao hoyuk=a*
beginning=LOC discussion be=F
''There will be a discussion at the beginning.'

where the locative suffix {=re} is added after the derivation of ɛ<tɔ>hɔp from ɛhɔp 'to begin' is over.

Infix can, therefore, be defined as a bound morpheme consisting of a minimal (single or double) phonological segment/s which splits apart a base and establishes a correspondence between the base (root or stem) and derived form. In the process, the derived form is morphologically complex as opposed to the source base, which is simple. Considering the discontinuity of the base as affected by splitting Yu (2003:2) defines infix as

> "...an overt continuous morph that appears with a derived discontinuous morph that appears in a continuous form independent of the infixed form, and the individual parts of this resultant discontinuous morph must not be continuous morphs themselves."

Infixation is an important morphological process operated in both nominal and verbal derivations in Munda. In adjectival constructions too, infixation sometimes plays an important role. Of all the Munda languages, both northern and southern[2], the system is more elaborate in Santali claiming five infixes <-n->, <-t->, <-m->, <-ɽ->, and <-p->, mostly as nominalizer, though at least two infixes <-p-> and <-t-> operate as verbalizer as well. Of these infixes, the most common one across Munda is <-n->, though in some languages like Korku and Ho where the infix is not a productive one[3], the vestiges show that the infix was once a productive derivational process in the languages concerned.

(4) Ko.*juʔ (jukh (V) rij 'to sweep') > ju<nu>ʔ 'broom' (Zide 2008:268). Compare Sa. jɔ<nɔ>k' 'broom' < jɔk' 'to sweep'; Mu. jo<no> 'broom' < jo' 'broom' (Bhaduri 1983:84)

(5) Ko. kaɽub- 'to cover' > ka<nu>ɽub 'lid' (Zide 2008:268).

(6) Ho. no<tu>m 'name' < *nom (?)

(7) Ho. ol 'to write, paint' > o<no>l 'coloured border of dhoti' (Burrows 1915:173)

The infixes <-t-> and <-m-> are common in the north Munda branch, except Korku, and in the south Munda branch, especially in Sora where both the infixes are attested and Gutob where <-m->, not <-t->, is attested. (In our data, no other language of the southern branch possesses these infixes.) Another infix <-p-> which is used as both a nominalizer and verbalizer is very common in the north Munda languages. In the Mon-Khmer group, especially in Khmer, Katu, Katu (Lao), Chrau, Sedang, Jeh, Surin Khmer, and Bahnar the most common infix in the nominal derivation is <-n->. Brian Migliazza (1998:64) reports that the infix <-n-> is also available in So, another Mon-Khmer language, as in pi:a:y 'to paddle' > pa:ni:a:y 'paddle'; cɔh 'to plant' > ci:nɔh 'plant'. The infix <-m-> is also attested in Khmer, as in Khmer sòm 'to beg' > s<m>òm 'beggar' (Jenner 1969:147). Besides the common infixes either in all the Munda languages or in the languages of the northern branch of Munda or in both Munda and Mon-Khmer (pan-Austroasiatic?), some infixes have also developed in individual Munda languages not shared by any other member of the group. <-ɽ-> in Santali, <-k-> and <-d-> in Korku, <-ər-> in Sora, and <-b-> in Remo are of this type. In our present study, we mainly focus on the infixes involved in nominal derivation and if any of the infixes straddles both the domains of nominal and verbal derivation that are taken care of.

As for the alignment (locus) of the infix, Blevins (2012) reports that it generally aligns to the beginning or end of the base. So far as the Munda languages (described here) are concerned, the infix is mostly inserted after the first C (as in Gta' in the monosyllabic base and in Sora in case of infix <-ər->, <-ət->, <-əm->, and <-ən->) and also after the first C followed by the vowel of the base in others. The infix, except Gta' in the monosyllabic base, and in languages where the infix has a fixed canonical shape with a vowel, get a vowel increment which is that of the base. In the disyllabic bases, sometimes it is found that the vowel of the second syllable follows the main part of the infix, that is, the consonant. This type of canonical shape and order of the infix partially matches with that of the Mon-Khmer languages. Normally, the vowel-initial bases insert the infix after the first vowel of the base, though insertion of the same before the base, that is, in the initial slot, is not uncommon. Such occurrences can be documented

from Sora and Gutob in Munda and some languages like Surin Khmer of the Mon-Khmer group.

Here data[4] have been drawn uniformly from all the major Munda languages and from some Mon-Khmer languages to review and elaborate the findings. Data have been drawn from both primary and secondary sources.[5]

2.0 Infixes developed in individual Munda languages

Sa. <-ɽ->, Korku <-k-> and <-d->, Sora <-ər->, Remo <-b->. Canonical shape [-cV-] except Sora and Korku in the case of <-k->. Locus: after the initial CV- of the base in Sa., Korku and Remo, and in Sora after the initial C-.

2.1 Santali <-ɽ->[6]

The infix is found only in Santali as an agentive nominalizer, as in the following examples:

(8) Sa. gɔm 'to name' > gɔ<ɽɔ>m 'grandparent or grandson / daughter'

(9) Sa. cɛ̃ 'to squeak' > cɛ̃<ɽɛ̃> 'bird'

2.2 Korku <-k-> and <-d->

Nagaraja (1999: 32) reports that in Korku, the process of infixation is least marked in nominal derivation. He mentions two infixes /-k-/ and /-dV-/ with the meaning of the result, and these two are not shared by any of the Munda languages, as in

(10) Ko. phaɽi 'to break/ split' > pha<k>ɽi 'bamboo wall'
(Nagaraja 1999:32)

(11) Ko. guɽi 'dry jowar plant' > gu<du>ɽi 'quilt' (Nagaraja 1999:32)[7]

2.3 Sora <-ər->

Ramamurthy (1986), Stampe and Donegan (2004:22-23), and Ghosh (2003: 224) report the existence of an infix /-ər-/[8], not shared by any Munda language, for deriving a noun with the meaning of locale or state as well as instrument.

2.3.1 Infix denoting locale or state

(12) So. baso ' to accommodate' (?) > b<Ar>aso 'accommodation'

(13) So. D<Ar>Akko 'place where things to be kept' < DAkko (?)

(14) So. g<Ar>An-na 'road' < gAn (?)

(15) So. g<Ar>en-na 'well' < gen? (Donegan & stampe 2004: 23)

(16) So. gob 'to sit' > g<ər>ob 'seat' (Ramamurti 1986:93); gɔb 'to sit' > g<Ar>ɔb 'seat'.

(17) So.ga 'to eat' > g<ər>a-ga 'plate for collecting food'

(18) So. pAsij 'child' > p<Ar>Asij 'childhood' (Donegan &Stampe 2004:23; also in Ghosh 2003)

(19) So. ba:l 'to burn'> b<ər>a:l-kul-ən 'a kiln' (bəra:l +kul-ən 'fireplace') (Ramamurti1986:46)

2.3.2 Infix with instrumental meaning

(20) So. Nam 'to catch' > N<Ar>am 'means of catching' [9]

(21) So. bɔj 'to sew' > b<Ar>ɔj-bɔj 'needle'

(22) So. g<Ar>en-Da 'that which serves to draw/ ladle water' < (?)

(23) So. b<Ar>Oj-kab 'needle' <?

(24) So. D<Ar>ab-Da 'a small dam for checking the run off of water' < Dab'to stop' + Da 'water' < Da'a 'water'.

(25) So. g<Ar>OsOD 'duster' < O. ghOs 'to clean, erase' (Donegan & Stampe 2004:22)

2.3.3 The infix can also used with derived stems, causative and reduplicated, to derive nouns

(26) So. babje (causative of baje 'to console') > b<ər>abje 'consolation' (Ramamurti 1986:46)

(27) So. gɨj 'to see' > g<ər>ɨj.gɨj-ən 'window' (Ramamurti 1986:92)

(28) So. ga 'to eat' > g<Ar>a.ga-si 'hand used for eating' (Donegan & Stampe 2004:22)

(29) So. tɨy 'to give' > t<ər>tɨy-tɨy 'vessel'

In vowel initial bases the infix is added before the first vowel, as in

(30) So. Ab-ga 'to feed' > <Ar>-Ab-ga 'spoon/plate' [10]

(31) So. Ab-suj (causative of suj 'to see') 'show'> <Ar>-Ab-suj-si 'index finger'

(32) So. agaD 'to sharpen' > <Ar>-agaD-ba 'whetting stone'

(33) So. AkOl 'to use a pruning hook to pluck fruit'> <Ar>-akOl 'hook for plucking fruit from trees' (Donegan & Stampe 2004:22)

2.4 Remo<-b->

Bhattacharya (1968: 70-71) and Ghosh (2003: 86) give only one example with infix <-b-> for deriving a noun denoting 'locality', as in

(34) Re. tuk' 'to weed' > tu<bu>k' 'earth' (Bhattacharya 1968:70-71)

3.0 Pan-Munda infixes <-t-> and <-m->

There are certain infixes in Munda that, though the evidence is not available from all the Munda languages, can be identified as pan-Munda if the languages like Santali, Mundari, and Ho, on the one hand, and Gutob and Sora, on the other, in which they are found, are regarded as representatives of north and south Munda respectively. The infixes in question are <-t-> and <-m->.

3.1 <-t->

The infix is found as a resultative nominalizer in Santali, Mundari, Ho, and Sora. It is also used as an agentive nominalizer in Santali and a nominalizer denoting locality in Mundari. In Santali, the infix is also found to derive a verb base from an existing base functioning as a noun and to derive nominals from the bases functioning as adjectives. In Santali, Mundari, and Ho the infix is realized as <-tV-> where V represents the base vowel. In the disyllabic base, however, the infix takes the vowel of the second syllable. In Sora, the infix gets a uniform phonological shape <-ət->.[11]

3.1.1 Infix as resultative nominalizer in Santali, Mundari, Ho, and Sora

(35) Sa. bɔr 'to fear' > bɔ<tɔ>r 'fear'

(36) Sa. rək' 'to sew' > rɔ<tɔ>k' 'seam'

(37) Sa. ɛhɔp' 'to begin' > ɛ<tɔ>hɔp' 'beginning'

(38) Sa. ɲum 'to name' > ɲu<tu>m 'name'

(39) Sa. ɔr 'to draw, pull' > ɔ<tɔ>r 'warp of a web'

(40) Mu. num 'to name' > nu<tu>m 'name' (Bhaduri 1983:134)

(41) Ho. *nom 'to name'> no<tu>m 'name' (Burrows 2015:172)

(42) So. ga:si 'to play' > g<ət>a:si 'game, sports' (Ramamurti 1986:93)[12]

The lexical form {notum} 'name' is found in Ho, though the base form {num} from which it is supposed to have been derived is not attested. Comparing its cognates in other Munda languages, especially Santali and Mundari it can be considered as a derived form although the root {num}' to name' is no longer in use in present-day Ho.

3.1.2 Infix as agentive nominalizer in Santali

(43) Sa. daram 'to oppose' > da<ta>ram 'councillor'

3.1.3 Infix as locational nominalizer in Mundari

(44) Mu. burum 'to lie down with folded legs' > bu<tu>rum 'resting place of an animal' (Hoffman 1950: 655)

3.1.4 The infix can also be inserted into a base functioning as an adjective to derive a noun in Santali

(45) Sa. jeleɲ 'long' > je<te>leɲ 'length'

3.1.5 Infix as verbalizer to derive a verb base from a noun is also found in Santali

(46) Sa. ɔr 'foundation' > ɔ<tɔ>r 'to demolish'[13]

As is found in Santali the infix can function as both nominalizer and verbalizer.

3.2 <-m->

The infix is found as an instrumental nominalizer to derive a noun from a verb base in Santali and as an agentive nominalizer for deriving nouns denoting specific relations and even pairs, from kin-terms in Santali and Gutob, and from a verb base in Gutob and Sora. While in Santali and Gutob the infix takes the vowel of the first syllable in monosyllabic bases, in Sora, the vowel of the infix is uniformly [ə]. The infix sometimes takes the vowel of the second syllable in Gutob as in 53, which is realized as [i] before [y] as in 52. The infix has parallels in some of the Mon-Khmer languages like Khmer (from old Khmer and modern Khmer) where the infix is found as an instrumental, agentive, and resultative nominalizer. Meng (2012:2) reports that "traditionally, the infix /-m-/ was known to play a key role in forming noun of agents and the infix /-n-/ also known to form noun of instruments". In her findings, she noted that "The interaction between the infix /-m-/ and roots creates many word classes of word formation such as Agentive nouns ... resultative nouns...instrumental nouns...verbs...and adjectives..." Drawing parallels from ancient inscriptions, she insists that it has existed since ancient times and continues in modern Khmer. While Meng identifies various functions of <-m-> including as an instrumental nominalizer, Jenner (1969:145), however, was of the view that "the functions of /-m-/ fall fairly clearly into two groups, one agentival, the other resultative". Jacob (1963:66-69)

considered <-m-> as a nominalized derivative referring to the agent of "base predication". Huffman (1967:93-94) also recognizes <-m-> as agentival nominalization, attribution, and predication. Jacob (1969:183-4) states that <-m-> denotes the agent of the base predication forms nouns and marks the causative. (Jenner 1969:145)

3.2.1 Infix as instrumental nominalizer in Santali

(47) Sa. lak' to peal, scrape' > la<ma>k' 'sliced fruit of jɔm laƫ used for scraping the ground or wall'. (Bo.2010:27)

(48) Sa. cɛt' 'to teach' > cɛ<mɛ>t' 'teaching material'

(49) Sa. sɛlɛt' 'to associate' > sɛ<m>lɛt' 'association' < sɛmɛlɛt'[14]

3.2.1.1 Infix as instrumental nominalizer in Khmer

(50) Khm. koh 'to play music' > kh<m>oh 'cymbal' (Meng 2012:2)

3.2.2 Infix as agentive nominalizer denoting specific relations in Santali and pairs in Gutob

(51) Sa. hɔn 'son' > hɔ<mɔ>n 'brother's offspring' when Ego is female'

(52) Gu. buyaŋ 'brother' > bu<mi>yaŋ 'two brothers'

(53) Gu. buloŋ 'sister' > bu<mo>loŋ 'two sisters'

3.2.3 Infix as agentive nominalizer in Gutob and Sora

(54) Gu. gul 'to bore' > gu<mu>l 'borer' (Griffith 2008: 651)

(55) So. g<əm>ul 'bug' (Rammurthy 1983:8) < (?); compare Gutob gul 'to bore'.

3.2.2.1 Infix as agentive nominalizer in Khmer[15]

(56) Khm. kăn 'to hold' > k<m>ăn 'holder'

(57) Khm. căm 'to watch' > c<m>ăm 'watchman'

(58) Khm. daə 'to walk' > t<m>aə 'pedestrian'

(59) Khm. baɲ 'to shoot' > p<m>aɲ 'one who shoots' (Jenner 1969:145-146)[16]

3.2.3 Infix as resultative nominalizer in Khmer, which function is not attested in Munda

(60) Khm. sòn 'to model, work clay'> s<m>òn ‘pottery’

(61) Khm. cùəɲ 'to trade'> c<m>ùəɲ 'business'

(62) Khm. lùəc 'to rob'> l<m>ùəc 'theft'

(63) Khm. rŏəh 'to scratch' > r<m>ŏəh 'itch' (Jenner 1969:147)

4.0 Pan-north Munda infix <-p->

The infix is operative in both nominal and verbal derivation in north Munda. The infix is found in many of the north Munda languages as a nominalizer to derive nouns denoting collectivity as in Santali, location as in Santali and Mundari, and degree or extent of adjectives and with "verbal reciprocal meaning" (AOH 2008) as in Ho. Reciprocal meaning is also not uncommon in Santali.

4.1 Infix <-p-> as nominalizer

4.1.1 Infix to derive noun denoting collectivity in Santali

(64) Sa. hɔn 'son' > hɔ<pɔ>n 'children'

(65) Sa. raj 'king' > ra<pa>j 'king and his retinue'

4.1.2 Infix as locative in Santali and Mundari,

(66) Sa. duɽup' 'to sit' > du<pu>ɽup' 'meeting place', as in dupuɽup' ʈə˜ɖi ‘meeting ground’

(67) Sa. gitic' 'to lie down' > gi<pi.tic' 'place of annual hunting'

(68) Mu. giti 'a place to sleep' of animal > gi<pi>ti 'a roost' (Hoffmann 1950:1458)

4.1.3 Infix to denote degree or extent of adjectives in Ho

(69) Ho. maraŋ 'big' > ma: <pa>raŋ 'very big' (AOH 2008:215)

Anderson, Osada, and Harrison report that in Ho nouns may also be derived with the infix <-p-> with the "verbal reciprocal meaning", although the source root is not always attested (Anderson, Osada, and Harrison, 2008: 213-14). The infix can also be found to derive a noun with reciprocal meaning in Santali.

(70) Ho. so<po>la 'reconciliation' (AOH 2008:213) < *sola (?)

(71) Ho. ku<pu>sar 'mutual enmity' (AOH 2008:214) < kusar (?)

(72) Ho. e<pe>ser 'counter claims of possession' (AOH 2008:214) < eser (?) cp. Mu. eser 'to possess'; Sa. eser 'to possess, occupy'

(73) Sa. bala 'marriage relation' > ba<p>la 'marriage

4.2 The infix acts as a regular verbalizer to derive reciprocal verb bases in north Munda

The infix acts as a verbaliser to derive reciprocal bases especially Santali, Mundari, Ho, and Korku, as in

(74) Sa. dal 'strike' > da<pa>l 'strike each other'

(75) Sa. rɔʈ 'speak'> rɔ<pɔ>ʈ 'quarrel with each other'

(76) Sa. ɛm 'give'> ɛ<pɛ>m 'give each other'

(77) Sa. ɔr 'pull'> ɔ<pɔ>r 'pull each other'

(78) Sa. tuɲ 'to pierce' > tu<pu>ɲ 'to pierce each other'

(79) Sa. tul 'to draw water' > tu<pu>l 'to draw for each other'

(80) Mu. dal 'strike'> da<pa>l 'strike each other' (Bhaduri 1983:37)

(81) Mu. or 'pull' > o<po>r 'pull each other' (Hoffmann 1950: 3118; Bhaduri 1983:183)

(82) Mu. tol 'bind'> to<po>l 'bind each other' (Bhaduri 1983:189)

(83) Mu. om 'give'> o<po>m 'give each other' (Bhaduri 1983:223) [Naguri dialect, in Hasada it is em > epem]

(84) Mu. eger 'abuse'> e<pe>ger 'quarrel with each other' (Bhaduri 1983:56)

(85) Mu. cal 'to spread disease' > ca<pa>l 'to impart disease to each other' (Hoffmann 1950:709-10)

(86) Mu. etoŋ 'to answer' > e<pe>toŋ 'to answer each other' (Hoffmann 1950:1335)

(87) Ho.* nel 'see'> ne<pe>l 'look at each other'

(88) Ho. goeʔ ' kill' > go<po>eʔ 'fight with each other'[17] (AOH2008: 214)

(89) Ko. munDa 'to beat' > mu<pu>nɖa 'to beat each other'

(90) Ko. sendra 'to walk'> se<pe>ndra 'walk together' (Nagaraja 1999:56)

(91) Ko. kumɑ 'to beat' > ku<pu>mɑ 'to beat each other'(Nagaraja 1999:319)

(92) Ko. men 'to say' > me<pe>n 'to consult' (Nagaraja 1999:325)

In Mundari, the infix can be inserted into a loan base also, which may be taken as proof for it being common and productive, as in

(93) Mu. cama 'to forgive' > ca<pa>ma 'to forgive each other' (Hoffmann 1950:718).[18]

4.3 Considering the application of the infix in major north Munda languages[19]

The application of the infix as a nominalizer and verbalizer can be considered common in north Munda. In the derived nominal forms found and given, there is the sense of 'collectivity', and the same sense is echoed through the reciprocal forms also.

5.0 Pan-Munda infix <-n-> and its Austroasiatic parallels

The infix <-n-> is pan-Munda, somewhere the phonological shape is constant with the addition of the root vowel following, thereby assuming the canonical shape <-nV->, in some cases, as in Sora, the infix takes the shape <-ən-> and in Gorum the infix takes shapes like <-an->, <-in->, or <-un->, thereby assuming the canonically <-Vn-> shape. In Gta', however, the infix is inserted after the initial consonant of the monosyllabic base without any vowel increment as the consonant clusters are permitted in Gta'. In disyllabic roots, normally the vowel of the first syllable repeats itself after the main part of the infix, that is, <-n->, although insertion of the vowel of the second syllable is not uncommon, as in Korku. The derived nouns denote either result of an action, instrumentality, or location. The extent or degree of the result can also be marked sometimes with the infix. In the Mon-Khmer group also, especially in Mon, Khmer, Bahnar, Katu, Katu (Lao), Chrau, Sedang, Jeh, and Surin Khmer of which the materials are available, the infix <-n-> with a vowel preceding or following or sometimes simply the consonant is inserted into the verb bases to derive a noun denoting result, instrumentality, or location.

5.1 Infix <-n-> as a resultative nominalizer

Derived nouns denoting 'result' of action in Munda:

(94) Sa. ɔl 'to write' > ɔ<nɔ>l 'writing, written piece'

(95) Sa. tɔl 'to tie' > tɔ<nɔ>l 'knot'

(96) Sa. ɔɽhɛ̃ 'to sing in praise' > ɔ<nɔ>ɽhɛ̃ 'eulogy'

(97) Sa. ɛʈɛt' 'to make fence'> ɛ<n>ɛʈɛt' 'partition/ fence'

(98) Mu. ol 'to write' > o<no>l 'writing/ written piece' (Bhaduri 1983:137-38)

(99) Mu. bai 'to make/ prepare' > ba<na>i ' creation' (Bhaduri 1983:19)

(100) Mu. be 'to spit' > be<ne> 'amount of spitting' (Hoffmann 1950:460) (Quantifier?)

(101) Mu. jul 'to burn' > ju<nu>l 'flame' (Bhaduri 1983:85)

(102) Mu. boto 'to gather for provision' > bo<no>to 'things gathered for provision' (Hoffmann 1950:627)

(103) Ho. tol 'to tie' > to<no>l 'bond'(AOH 2008:213)

(104) Ho. chaba 'to finish' > cha<na>b 'end' (Burrows 1915:154) < cha<na>ba.

(105) Ho. em 'give' > e<ne>m 'gift' (AOH 2008:213)

(106) Ho. ol 'to write, paint' > o<no>l 'coloured border of dhoti' (Burrows 1915:172)

(107) Ko. kaɽub 'to cover' > ka<nu>ɽub 'cover' (Zide 2008:268)

(108) Kha. kuʔj 'to dance' > ku<nu>ʔj 'dance'

(109) Kha. bel 'to spread (mat)' > be<ne>l 'bedding' (Peterson 2008: 452)

(110) So. gay 'to dig' > g<ən>ay 'tuber'[20]

(111) So. *gur 'to rain' > g<ən>ur 'rain'[21]

(112) So. ɖul 'to compensate' > ɖ<ən>[r]ul 'compensation' (Donegan &Stampe 2004:71)

(113) So. gu 'to plant' > g<ən>u 'planting' (Donegan & Stampe 2004: 80)

(114) So. gam 'to speak/ narrate' > g<ən>um 'narration' (Donegan & Stampe 2004:85)

(115) Go. ab 'to husk' > <an>ab 'paddy husk'[22] (Donegan & Stampe 2004:3)

(116) Go. al 'to thatch' > <an>al 'thatch' (Donegan & Stampe 2004:12)

(117) Go. galbaʔ nu 'to tie a turban on head' > g<in>albaʔ 'turban'[23] (Donegan & Stampe 2004:59)

(118) Go. gomteʔn 'to dream' > g<un>omti 'dream' (Donegan & Stampe 2004:65)

(119) Gu. ab 'to scoop out' > <an>ab 'husk'[24]

(120) Gu. biʔtir 'to spit' > b<in>iʔtir 'saliva' (Donegan & Stampe 2004:9)

(121) Gu. baj 'to draw, decorate' > b<un>aj 'writing, embroidery' (Donegan & Stampe 2004:7)

(122) Re.tap' 'to cover' > ta<na>p' 'shade'

(120) Re. tupak' 'to bundle' > tu<nu>pak' 'load, bundle'

(121) Re. ɖai 'to climb, ride' > ɖa<na>i 'ghat on the hills' (Sahu, Samantaray & Patel 1993:59)

(122) Re. ɖɔk' 'to make a trap' > ɖɔ<nɔ>k' 'trap'

(123) Gt. ciaʔ 'to borrow'> c<n>iaʔ 'debt'(L.)

(124) Gt. cuaʔ 'to collect donation' > c<n>uaʔ 'donation'

(125) Gt. baʔlir 'to talk, converse' > b<n>aʔlir 'conversation'

(126) Gt. cog 'to put on ornament' > c<n>og 'ornament'

(127) Gt. coʔ 'to sweep' > c<n>oʔ 'broom' (Panda 1989:14-16)

5.1.2 Derived nouns denoting 'result' of action in Mon-Khmer

In the Mon-Khmer group also the infix <-n-> with a vowel preceding or following or sometimes simply the consonant is inserted into the verb bases to derive nouns denoting result, that which is made by the action of the verb, as in

(128) O.M. gruŋ 'to laugh' > g<in>ruŋ 'laughter'

(129) O.M. p'ār 'to put into practice' > p<un>'ār 'conduct' (Jacob 1963:19)

(130) O.Khm. sre 'rice land' > s<an>re 'rice-field (quantifier) (Jacob 1960:351-68; 1963:24)

(131) M.Khm. kdap 'to grasp in the hands' > k<ɔn>dap 'sheaf' (Jacob 1963:24)

(132) Bah. pah 'to split' > p<o'n>ah 'split bamboo'

(133) Bah. dan 'to place as target' > d<o'n>an 'a target'

(134) Bah. tār ‘to weave’ > t<o’n>ār ‘woven bamboo’

(135) Bah. bât ‘to make a dam’ > b<o’n>ât ‘a dam’

(136) Bah. kao ‘to make a wedge’ > k<o’n>ao ‘a wedge’

(137) Bah. chĕng ‘to make a partition’ > ch<o’n>ĕng ‘partition’ (Banker 1964:101)

(138) Ka. cha 'to name' > ch<an>a 'name'

(139) Ka. gi 'to plan' > g<an>i 'a plan'

(140) Ka. chiam ‘to feed’ > ch<an>iam ‘food given’

(141) Ka. chóór ‘to make groove on crossbow’ > ch<an>óór ‘groove on crossbow’

(142) Ka. chuul ‘to make sound’ > ch<an>uul ‘sound of animals, music’

(143) Ka. dóók ‘to name’ > d<an>óók ‘name’

(144) Ka. klâm ‘to urinate’ > k<a>lâm ‘urine’[25] (Costello 1966:63-86)

(145) Chr. poq 'to roll' > p<an>oq[26] 'one roll (cigarette)’

(146) Chr. tuŋ 'to carry on shoulder' > t<an>uŋ 'one load (of firewood)’ (Thomas 1969:104)

(147) Sed. kang 'thing for fencing'[27]> k<on>ang 'fence'

(148) Sed. péng ‘thing for shooting’ > m<on>éng ‘crossbow’

(149) Sed. chia ‘thing for digging’ > ch<on>ia ‘hoe’

(150) Sed. dêa ‘thing for adding water to water jug’ > d<on>êa ‘measuring stick’

(151) Sed. krum 'thing for fencing plant' > k<on>rum 'fence around plant' (Smith 1969:124)

(152) Je.[28] tuy 'to carry' > t<an>uy 'load over shoulder'

(153) Je. pra 'to spread out' > m<ad>ra 'trellis'

(154) Je. prăt ‘to switch’ > p<ad>răt ‘a switch’

(155) Je. buh ‘to hammer’ > m<an>uh ‘a hammer’ (Gradin 1976:25-42)

(156) Su.Khm. kuar 'to mix together' > k<n>uar 'a mixture'

(157) Su.Khm. saap 'to sprout' > s<n>aap 'seedlings'

(158) Su.Khm. kuur 'to draw' > k<n>uur 'wavy line'

(159) Su.Khm. teʔ 'to find fault' > t<n>eʔ 'fault' (Thomas 1990: 91)

(160) Ka.(Lao). kui 'to carry on back' > k<an>ui 'something carried on back'

(161) Ka.(Lao). tôl 'to put post in ' > t<an>ôl 'post'

(162) Ka.(Lao). pó 'to dream' > p<an>ó 'a dream'

(163) Ka.(Lao). kuôl ‘to have resources, strength’ > k<an>uôl ‘resource, strength’

(164) Ka.(Lao). kloos 'to exchange' > k<a>loos 'an exchange'

(165) Ka.(Lao). klam 'make tree shrine' > k<a>lam 'tree shrine'[29] (Costello 1998:31-42)

5.2 Infix <-n-> as an instrumental nominalizer

Infix as an instrumental nominalizer in Munda: The same infix <-n-> is also found to derive a noun denoting instrument in most of the Munda languages, as in

(166) Sa. jɔk' 'to sweep' > jɔ<nɔ>k' 'broom'

(167) Sa. bak' 'to hook' > ba<na>k' 'hook'

(168) Mu. tagoeʔ 'to chew' > ta<na>goeʔ 'the molar teeth' (Osada 2008:116)

(169) Mu. jo' to sweep' > jo<no> 'broom'

(170) Mu. goŋ 'to give a girl in marriage' > go<no>ŋ 'bride price' (Bhaduri 1983:62)

(171) Ho. joʔ 'to sweep' > jo<no>ʔ 'broom'

(172) Ho. rapiḍ 'to wink, blink' > ra<na>piḍ 'eyelid' (AOH 2008:213)

(173) Ho. soro 'to shut' > so<no>ro 'the bar across a door to keep it shut' (Burrows 1915)

(174) Ko. jukhṛij 'to sweep' > ju<nu>ʔ 'broom' (Zide 2008: 268)

(175) Kha. si 'to plow' > si<ni> 'plow'

(176) Kha. ɖeʔj 'to chop' > ɖe<ne>ʔj 'hatchet' (Peterson 2008:452)

(177) Ju. jɔg 'to sweep' > j<ən>ɔg 'broom'

(178) So. gOD 'to sharpen' > g<An>OD 'instrument used for sharpening knife' (Donegan & Stampe 2004:87)

(179) So. jO 'to sweep' > j<An>O 'broom' (Donegan & Stampe 2004:101) [ɔ is transcribed as O in D & S]

(180) Go. kuŋ 'to shave' > k<an>uŋ 'razor' (Donegan & Stampe 2004)

(181) Go. raj 'to comb' > r<in>ay 'comb'

(182) Go. jujɔʔ 'to wipe' > j<ɔn>ɔ 'broom' (Anderson & Rau 2008:396)

(183) Gu. sui 'to plow' > su<ne>i 'plow'

(184) Gu. noŋ 'to put a yoke on' > n<in>oŋ 'yoke'

(185) Gu. gir 'to fish' > g<in>ir 'fishnet'

(186) Gu. peɖ 'to blow a flute' > p<in>eʔ 'flute'

(187) Gu. boʔ/ɖ 'to hang/suspend' > b<un>oʔ 'ladder'

(188) Re. suk' to sweep with soft broom' > su<nu>k' 'soft broom'

(189) Re. ɔkser 'to hang' > ɔ<nɔ>kser 'hanger'

(190) Re. ɖɔk' 'to make a trap' > ɖɔ<nɔ>k' 'trap'

(191) Re. gurak' 'to spin thread' > gu<nu>rak' 'spindle for spinning thread'

(192) Re. sug/suk'bɔ' 'to comb hair' > su<nu>gbɔ' 'comb'

(193) Gt. coʔ 'to sweep' > c<n>oʔ 'broom'

(194) Gt. Doʔ 'to trap' >D<n>oʔ 'trap'

(195) Gt. oʔriaʔ 'to wind thread in a specific frame' > o<no>ʔriaʔ 'bobbin'

(196) Gt. caliʔ 'to taste' > c<n>aliʔ 'wine tasting stick'

(197) Gt. oro 'to hang' > o<no>ro 'rope hanger'

5.2.2 Infix as an instrumental nominalizer in Mon-Khmer

In the Mon-Khmer languages also, the infix has the same function, that is, the derived noun denotes ‘instrument’, that which is used to perform an action, as in

(198) Khm. chok 'to cover' > ch<n>ok 'cork, lid'

(199) Khm. khoj 'to rest on a pillow' > kh<n>oj 'pillow'

(200) Khm. khar 'to spin' > kh<n>ar 'spindle' (Meng 2012:1-7)

(201) Ka. gap 'to cut with scissors' > g<an>ap 'scissors'

(202) Ka. panh 'to shoot' > p<an>anh 'crossbow'

(203) Ka. ten 'to hammer' > t<an>en 'hammer'

(204) Ka. glâk ‘to carry by two people’ > g<a>lâk ‘pole for carrying’[30]

(205) Ka. pruung ‘to blow fire’ > p<a>ruung ‘pipe to blow fire’

The infix is active in modern Katu as in

(206) Ka. chui ‘to erase’ > ch<an>ui ‘eraser’

It may also be added to the loanwords from Vietnamese, as in

(207) Ka. geng ‘to carry with pole’ > g<an>eng ‘carrying pole’ (Costello 1966:63-86)

(208) Chr. goch 'to lasso' > g<an>och 'lasso'

(209) Chr. păr 'to fly' > p<an>ăr 'wings' (Thomas 1969:104)

(210) Chr. chia 'thing for digging' > ch<on>ia 'hoe'

(211) Chr. dêa 'thing for adding water to wine jug' > d<on>êa 'measuring stick (to know when to add water)

(212) Chr. péng ‘thing for shooting’ > m<on>éng ‘crossbow’

(213) Je. buh 'to hammer' > m<an>uh 'hammer'[31]

(214) Je. gap 'to hold (with slit stick) > k<an>ap 'a slit stick'

(215) Je. prăt ‘to switch’ > p<ad>răt ‘a switch’ (Gardin 1976:26-27)

(216) Su.Khm. ket 'to wipe anus' > kh<n>et 'wiper'

(217) Su.Khm. set 'to comb' > s<n>et 'comb'

(218) Su.Khm. ʔar ‘to saw’ > ʔ<an>ar ‘a saw’ (Thomas 1990:85-98)

(219) Bah. pai 'to cook' > p<o'n>ai 'a stirring stick'

(220) Bah. par 'to fly' > p<o'n>ar 'wing' (Banker 1964:101)

5.3 Infix <-n-> as a locational nominalizer

The derived nouns in Munda with the infix /-n-/ sometimes denote location, that is, the place where the action is performed.

(221) Sa. dul 'to pour' > du<nu>l 'the junction of two rivers/ streams where water falls'

(222) Sa. rakap' 'to ascend'> ra<na>kap' 'ascending slope'

(223) Mu. rakab' 'to rise' > ra<na>kab' 'a rising ground' (Bhaduri 1983:156)

(224) Mu. eteʔ 'to begin' > e<ne>ʈeʔ 'origin' (Osada 1992:62)

(225) Ho. a:du 'to descend' > a<na:>du 'steep descent'

(226) Ho. rakab 'to climb' > ra<na>kab 'steep ascent' (Burrows 1915)

(227) Kha. raʔb 'to bury' > ra<na>ʔb-raʔb 'burial ground, grave' (Pinnow:1963)

(228) So. gu 'to plant' > g<ən>u 'garden, grove'

(229) So. ɖakko 'to keep' > ɖ<ən>akko 'place where something is to be kept' (Donegan & Stampe 2004)[32]

(230) Go. al 'to thatch' ><an>al 'thatch'[33]

(231) Go. raŋ 'to keep' > r<in>aŋ 'wooden platform for keeping things' (Donegan & Stampe 2004 op.cit)

(232) Re. ɖai 'to climb, ride' > ɖa<na>i 'ghat on the hill'

(233) Re. ruk' 'to open door' > ru<nu>k' 'courtyard' (Bhattacharya 1968: 59, 115)

(234) Gt. ugtur 'to make a fireplace' > u<nu>gtur 'oven made of three stones' (Panda 1987:7)

(235) Gt. waʔɽɛŋ 'to dance to drum beats' >wa<na>ʔɽɛŋ 'dancing ground' [34]

The infix <-n-> can perform the function of a locational nominalizer in the Mon-Khmer languages also, as in

(236) Khm. khoŋ 'to reside' > kh<n>oŋ 'residence' (Meng 2012:4)

(237) Khm. pɛn 'to sit upon the crossed legs' > p<n>ɛn 'the horizontal plane formed by sitting upon the crossed legs' (Jenner 1969:53)

(238) Ka. bêch 'to sleep' > b<an>êch 'bed'

(239) Ka. to'o't 'to sit' > t<an.o'o't 'stool'(Costello 1966:65)

(240) Chr. paŋ 'wall in' > p<an>aŋ 'room'

(241) Chr. gap 'to dam up' > g<an>ap 'dam' (Thomas 1969:104)

(242) Je. pring 'to go' > p<ad>ring 'road'

(243) Je. tìeng 'to dry in sun' > t<an>ìeng 'sunny place' (Gradin 1976:30)

(244) Su. Khm. daʔ 'to place' > t<n>aʔ 'place where placed on trip' (Thomas 1990:88)

(245) Ka. (Lao). tôl 'to put post in' > t<an>ôl 'post'

(246) Ka. (Lao). trooq 'to make enclosure' > t<a>rooq 'enclosure' (Costello 1998:39-40)

(247) Bah. puŋ 'to wallow' > p<o'n>uŋ 'wallowing hole' (Banker 1964:102)

Some points need to be commented upon here; one, the main part of the infix is <-n-> irrespective of being augmented by preceding or following vowel or null (except for Katu and Katu (Laos) where it is reduced to <-a-> when it is inserted into clusters); two, in three important languages Gta' of the Munda group, Old Mon, and Khmer with Surin Khmer (a dialect of Khmer) of the Mon-Khmer group, the infix is inserted after the initial consonants without any vowel increment. In these languages consonant cluster is permissible. While it is a regular practice in Gta' in monosyllabic bases, it is phonologically conditioned in Mon-Khmer. Three, syntactically and semantically the infix functions as a nominalizer.

5.4 Extension of the function of <-n-> in some languages

Nominalizing aside, the infix has acquired other functions as well. In Remo, the infix <-n-> is also used as a verbalizer in deriving a reciprocal stem, as in

(248) Re. buk' 'to beat' > bu<nu>k' 'to fight with each other'

(249) Re. rap' 'to pull' > ra<na>p' 'to pull one another'. (Bhattacharya 1968:95-6, 113)

In the same language, the infix is used to derive inclusive kin-terms and also to form dual of nouns, as in

(250) Re. tuna 'younger sister' > tu<nu>na 'younger brother and sister'

(251) Re. biyaŋ 'younger brother' > bi<ni>yaŋ 'two brothers'

(252) Re. bailɔk' 'friend' > ba<na>ilɔk' 'two friends'

(253) Re. giriŋ 'wife's younger brother' > gi<nd>riŋ 'I and my wife's younger brother'. [with homorganic stop]

It can act upon numerals to derive inclusive nominals in Santali, as in

(254) Sa. bar 'two' > ba<na>r 'both'

(255) Sa. pɛ 'three' > pɛ<nɛ> 'all three'.

(256) Sa. pon 'four' > po<no>n 'all four'

It may denote other meanings too, like, expressing degree or extent in Mundari, as in

(257) Mu. bul 'to intoxicate' > bu<nu>l 'degree of intoxication'. (Hoffmann 1950:644)

In Gta', the infix may also be inserted into a nominal base to derive further noun, as in

(258) Gt. kala 'moment' (L) > k<n>ala 'a particular point of time'

It is sometimes found to be used as an infix for adjectivization, as in Chrau,

(259) Chr. vŏh 'to know' > v<an>ŏh 'wise' (Thomas 1969:105)

Loss of earlier affixes and subsuming different functions under the remaining affixes may be the cause behind assuming several meanings of the same infix. It is doubtless that the infix <-n-> is common in both Munda and Mon-Khmer, though the basic functions like resultative, instrumental, and locative nominalizer are not uniformly attested in all the languages of the groups. It also happens that the derived forms with the infix are found as vestiges although the source base is no longer traceable or in use, as in

(260) So. g<ən>ur 'rain' < *gur, cp. Kharia gur 'to rain',

(261) Go. k<an>uŋ 'razor' < *kuŋ, cp. Sora koŋ 'to shave',

(262) Kha. g<an>a?ɖ 'sickle' < *ga?ɖ, cp. Sora ga?ɖ 'to cut'.

It also happens that in some languages the derived form is used both with the root meaning and the derived meaning. Dorothy M Thomas (1969: 90-107), while discussing the Chrau Affixes, has given classic examples of that situation.

6.0 Regarding infix /-m-/ in Munda and Khmer

It is seen earlier in the chapter that the nouns denoting 'instrument' are derived with the infix <-m->, and at least one agentive noun (kin-term) is derived along with it. In Gutob, one example with agentive meaning is found to be derived with the infix. Possibly these are not stray instances and have some historical significance. In Khmer too, the infix is used as an agentive, instrumental, and resultative nominalizer. Though the meaning of 'result' is not attested in Munda, the latter resembles Khmer with the meaning of 'agent' and 'instrument'. In that sense, the infix <-m-> may also be postulated as a common Munda-Mon-Khmer infix though it is premature to prove at this stage with minimum examples from Munda that it is derived from proto-Austroasiatic <-m->. Further investigation may help us prove the hypothesis with greater certainty.

7.0 Conclusion

Infixation so far discussed is a class changing device at the derivational level. In cases where the infix does not operate as a class changer, as in four examples of Remo, one example of Gta' discussed in section 5.4 under other functions of <-n->, in Sedang where nominals are derived from noun base from <-n->, and Gutob where <-m-> derives natural pairs from kin-terms, it virtually entails a change within the same class. The whole phenomena of infixation have been categorized in four parts: one, there are certain infixes, like <-ɽ->, <-k->, <-d->, <-ər->, and <-b-> in some Munda languages which have developed in the individual languages like Santali, Korku, Sora, and Remo; two, there are some infixes like <-p-> which are shared by the major languages of the north Munda like Santali, Mundari, Ho, and Korku; three, there are two infixes like <-t->

and <-m-> shared by some (not all) languages of both the branches, north and the south, like Santali, Mundari, Ho, Gutob, and Sora. While <-t-> is shared among the Munda languages, <-m-> is shared by Khmer outside Munda; and four, there is one infix, <-n->, which is shared by all the languages of Munda and more than ten languages of the Mon-Khmer group. Historically speaking, in the first case, it may be assumed that the infixes developed in the individual languages well after the individual languages grew. <-p ->, which is commonly shared among the North Munda languages, may be postulated in the proto-north Munda. Similarly *<-t-> may be postulated in the proto-Munda stage as some languages of both the groups share it. <-m-> as shared by some languages of both north and south Munda and most importantly by Khmer may be postulated in the proto stage of both the groups. The most common infix in both Munda and Mon-Khmer is <-n-> which with utmost certainty can be proved to be belonging to the proto-stage of both the groups. The infix <-p-> which is common as a reciprocal formation in north Munda is not attested in South Munda. Pinnow (1960) from his comparison of Munda and Khmer-Aslian tried to establish <-p-> as a proto-Austroasiatic infix. Anderson (2008) also calls the infix <-n-> as pan-Munda, pan-Austroasiatic. Technically speaking, our methodology is not a historical-comparative one; our main intention is to study nominal infixes from a pan-Munda perspective and locate parallels in the Mon-Khmer languages. There are several historical-comparative studies of the Mon-Khmer infixes, especially <-n-> and <-m-> and some of those have also made certain reconstructions. We have not referred to those studies or reconstructions because of the difference in focus.

Abbreviations

By language names:
Bah.= Bahnar; Chr.= Chrau; Go.= Gorum; Gu.= Gutob; Gt.= Gta'; Ho.=Ho; Je.=Jeh; Ju.= Juang; Ka.= Katu; Ka.(Lao)= Katu (Lao); Kha.= Kharia; Khm.= Khmer; Ko.=Korku; MKhm.= Modern Khmer; Mu. Mundari; OKhm.= Old Khmer; OM= Old Mon; Re.=Remo; Sa. = Santali; Sed.=Sedang; So.= Sora; Su.Khm.= Surin Khmer.

Other abbreviations:
A= applicative; AOH= Anderson, Osada and Harrison; Bo.= Bodding; F= finite marker; L= loan; LOC= locative; O= object; pl= plural; S= subject; sg.= singular;

Notes

[1] Following Leipzig Glossing Rules the infix is put within <>.
[2] The Munda group of languages is broadly divided into two branches, viz., northern and southern. Santali, Mundari, Ho, Birhor, Korku, Korwa, Asuri, Turi belong to the northern group, spoken in Jharkhand, Bihar, Odisha, West Bengal, Chattisgarh, Madhya Pradesh, and Maharashtra. Kharia, Juang, Sora, Gorum, Remo, Gutob, and Gta' belong to the southern group spoken mainly in Jharkhand, Odisha, and Andhra Pradesh.
[3] Nagaraja (1999), Zide (2008), and Anderson, Osada, Harrison (2008) report that the infix <-n-> is not productive or a weakly developed one in Korku and Ho, respectively.
[4] Data are arranged in numerical order followed by language name in abbreviation.
[5] Primary sources are the author's published materials from 1994 and 2003, and unpublished materials from the field notes of Remo, Gutob, Gta' and Gorum collected in 2009-2011. Where the source is not given after the data, it means that those are drawn from primary sources. The secondary data are, however, drawn from various published materials shown in parenthesis after the data.
[6] This infix as nominalizer may be found in some north Munda languages, though for a paucity of data we cannot exemplify from other languages.
[7] While the infix <-k-> inserted into the base without any vowel increment as in 10, in 11, however, the infix<-d-> takes a vowel increment which is the vowel of the first syllable.
[8] The infix <-ər-> is transcribed as <-Ar-> by Donegan & Stampe (2004) and Ghosh (2003) while Ramamurti (1986) transcribes it as [ə].
[9] The palatal nasal [ɲ] is transcribed as [N] in Ghosh (2003).
[10] This type of adding infix at the beginning of a base can also be had in some Mon-Khmer languages, especially in Jeh Donegan and Stampe identify this <Ar> as both prefix and infix (Donegan & Stampe 2004:22)
[11] <-At-> in Donegan & Stampe (2004) and Ghosh (2003).
[12] Stampe and Donegan in their online dictionary of Sora (2004) have not mentioned this infix and Ghosh 2003: 224 transcribes the vowel of the infix as [A] as in gAtasi 'play'< gasi 'to play'.
[13] For detailed discussion on Santali derivation see Ghosh 1994: 20-35 & 2008:50-52).
[14] sɛmɛlɛt' > sɛmlɛt' due to vowel reduction as the maximum limit of word structure in Santali, maybe in other Munda languages too, disyllabic.
[15] For detailed study in Khmer<-m-> please see Meng (2012:1)
[16] The infix is inserted into the base after the initial consonant without any vowel increment. The initial consonant of the base is realized as [-voice] before <-m->.
[17] The derived forms in 87 and 88 are also used as nouns in the sense of 'mirror' and 'fight' respectively.
[18] The base {cama} is borrowed from Sadani {cama} 'to pardon', compare Skt., Hindi *kshama*:.
[19] By 'major' we mean those languages are spoken by a large number of speakers and which are well documented as opposed to the minor ones, like Asuri, Birhor,

Turi and others which are spoken by a small number of speakers and which are also under-documented.

[20] The infix was transcribed as <-An-> in Ghosh (2003) Here transcription is changed to <-ən-> as per fieldnotes 2009.

[21] Although the root <gur> is not attested in Sora it can be posited as Munda base, compare Kharia {gur} 'fall of rain'.

[22] It is sometimes assumed that in vowel-initial bases locus of the infix with a vowel preceding is the initial position. This is the position that inspires Donegan & Stampe (2004) to describe the [-Ar-], <-Am->, and [-An-] as prefix and infix. On the basis of the data from Gorum and Gutob in examples 115-116 and 119 an alternative scanning is proposed, that is, the infix is inserted after the initial vowel of the base.

[23] The context in which the infix assumes its phonological shape may tentatively be located in the following consonant, [-in-] before alveolo-palatal [l] and [-un-] before plosive [m], as in 117 and 118.

[24] Same as note 23.

[25] In bases with initial cluster the <-n-> is phonologically reduced to <-a->, elsewhere realised as <-an->.

[26] In Chrau <-an-> occurs in monosyllabic base only and follows the initial consonant.

[27] In Sedang /-on-/ derives further nouns from noun base.

[28] The infix is normally <-an-> in Jeh. It also changes according to the context, as in 153-154 where <-an-> changes to <-ad-> as <nr> cluster is not permissible in Jeh. It may also cause a change in the initial consonant of the base, as in 153 and 155, as <-ad-> precludes changes in the consonant before it and ba- does not occur as a presyllable in the language respectively.

[29] The infix <-an-> in Katu (Lao) alternates with <-a-> when it is inserted into verb bases with initial cluster, as in 164 and 165.

[30] The infix <-n-> is realised as <-a-> before [r] and [l] as -nr- and -nl- combinations are not permitted in Katu.

[31] The same condition holds as in note 28.

[32] Data is taken from Donegan & Stampe, re-transcripsion is that of the author.

[33] Same proposal as in note 22.

[34] 2003 data re-transcribed.

References

Anderson, Gregory D. S, ed. *The Munda Languages*. London & New York: Routledge, 2008.

Anderson, Gregory D.S., Toshiki Osada, and K. David Harrison. “Ho and the other Kherwarian Languages.” In *Munda Languages,* edited by Gregory D.S. Anderson, 195-255, London & Newyork: Routledge, 2008.

Anderson, Gregory D. S, and Felix Rau. "Gorum." In *Munda Languages,* edited by Gregory D.S. Anderson, 381-433, London & Newyork: Routledge, 2008.

Banker, Elizabeth M. "Bahnar Affixation." *MKS* 1 (1964): 99-117.

Bhaduri, Manindra Bhusan. *A Mundari English Dictionary*. New Delhi: Northern Book Centre, 1983.

Bhattacharya, Sudhibhushan. *A Bonda Dictionary*. Poona: Deccan College, 1968.

Blevins, Juliette, "Infixation" http://www.julietteblevins.ws.gc.cunny.edu/files/2016/10/Blevins 2012-infixationrev1.pdf dated 29 March 2012.

Bodding, P.O. *A Santhal Dictionary*. New Delhi: Gyan Publishing House, 2010.

Burrows, L. *The Grammar of the Ho Language*. Delhi: Cosmos, 1915.

Costello, Nancy A. "Affixes in Katu." MKS 2 (1966): 63-86.

—. "Affixes in Katu of the Lao P.D.R." MKS 28 (1998):31-42.

Donegan, Patricia & David Stampe. "Austroasiatic On-line Dictionary", www.ling.hawaii.edu/austroasiatic/AA/Munda/Dictionaries/korku/Gorum/Sora, 2004.

Ghosh, Arun. *Santali: A Look into Santal Morphology*. New Delhi: Gyan Books, 1994

—. *An Ethnolinguistic Profile of Eastern India: A Case of South Orissa.* Burdwan: University of Burdwan, 2003.

—. "Santali." In *Munda Languages,* edited by Gregory D.S. Anderson, 11-98, London & Newyork: Routledge, 2008.

Gradin, Dwight. "Word Affixation in Jeh." *MKS* 5(1976): 25-42.

Griffiths, Arlo. "Gutob." In *Munda Languages,* edited by Gregory D.S. Anderson, 633-681, London & Newyork: Routledge, 2008.

Hoffmann, Rev. J. *Encyclopaedia Mundarica*. Patna: Government Press, 1950.

Huffman, Franklin Eugene. "An Outline of Cambodian Grammar." Ph.D. dissertation, Cornell University, 1967.

Jacob, Judith M. "Prefixation and Infixation in Old Mon, Old Khmer and Modern Khmer." In *Linguistic Comparison in Southeast Asia and Pacific* edited by H. L. Shorto, 62-70. London: SOAS, 1963.

—. *Introduction to Cambodian* [with magnetic tapes] London: Oxford, 1968

—. "The structure of the word in Old Khmer." *BSOAS* 23, no.2 (1960):351-368.

Migliazza, Brian, "A Grammar of So-- A Mon-Khmer language of Northeast Thailand." Ph.D. thesis. Mahidol University, 1998.

Meng, Vong. “Infixes /-M-/ and /-N-/ in Khmer Languages.” *Cambodia Studies Conference: Imaging Cambodia*, Centre for Southeast Asian Studies at Northern Illinois University. 13-16 September 2012.

Nagaraja, K.S. *Korku Language*. Tokyo: Tokyo University of Foreign Studies, 1999.

Osada, Toshiki. “Mundari.” In *Munda Languages,* edited by Gregory D.S. Anderson,99-164, London & Newyork: Routledge,

Panda, Pramoda Kumari. *Didayi*. Bhubaneswar: Academy of Tribal Dialects and Culture, 1989.

Peterson John. *A Grammar of Kharia*. Leiden: Brill, 2011

—. “Kharia.” In *Munda Languages,* edited by Gregory D.S. Anderson, 434-506, London & New York: Routledge, 2008.

Ramamurti, G.V. *Sora-English Dictionary*. New Delhi: Mittal Publication, 1986.

Sahu, Kalindicharan, Saktiprasanna Samantaray and Paramananda Patel. *Bonda*. Bhubaneswar: Academy of Tribal Dialects and Culture, 1993.

Sey, Yoy. “Some Old Khmer Affixation.” *MKS* 5(1976): 85-95.

Smith, Keneth D. “Sedang Affixation.” *MKS* 3(1969): 108-129.

Stump, Gregory. "Word formation and Inflectional morphology." In *Handbook of Word Formation* edited by Pavol Stekauer and Rochelle Leiber, 49-72. Dordrecht: Springer, 2005.

Thomas, Dorothy M. “Chrau Affixes.” MKS 3(1969): 90-107.

—. “The Instrument/ Locative and Goal Affix -N- in Surin Khmer.” MKS 16-17 (1990):85-98.

Yu, Alan Chi Lun, “The Morphology and Phonology of Infixation.” Ph.D. Dissertation, University of California, Berkeley, 2003

Zide, Norman H. “Korku.” In *Munda Languages,* edited by Gregory D.S. Anderson, 256-298, London & New York: Routledge, 2008.

CHAPTER THREE

AGREEMENT REVERSAL IN MUNDA LANGUAGES: AN INTERPLAY OF FUNCTIONAL/THEMATIC AND SYNTACTIC CRITERIA

KĀRUMŪRI V. SUBBĀRĀO, AND MARTIN EVERAERT

1. Introduction*

1.1 Non-Nominative Subjects

Though Non-Nominative Subjects (NNS hereafter) are generally considered as a diagnostic feature for 'India as a Linguistic Area', it is significant that many, but not all[1], Tibeto-Burman languages and all Mon-Khmer languages do not have the NNS construction. NNS in South Asian languages have been discussed in Verma & Mohanan (1990), Pandharipande (1979), Mohanan & Mohanan (1990), Bhaskararao & Subbarao (2004), Subbarao & Bhaskararao (2004), Subbarao (2012), and Butt (2013) amongst others. However, there is no work done on the nature of NNS in any Munda language to the best of our knowledge. This paper will address that lacuna.

Like in Icelandic, German, Russian, and Dravidian and Indo-Aryan languages, the NNS, though obliquely case-marked, retains its property of a subject with regard to (i) being an antecedent to an anaphor, (ii) being the controller of the notional subject, generally labeled as PRO in non-finite object complement clauses, (iii) being the controller of the embedded subject in a conjunctive participial construction, etc. Except for a couple of languages of the Indo-Aryan family, the NNS does not trigger agreement on the predicate (Subbarao 2012). The predicate (verb or

adjective) in a non-nominative subject construction is [-transitive], and it is non-volitional. Since it is [-transitive], the predicate cannot assign the accusative case to the theme or patient. Hence, the theme/patient is always the nominative case- marked.[2]

Subbarao & Bhaskararao (2004) and Subbarao (2012) have shown that in Indo-Aryan and Dravidian languages there are some specific functional domains (types of predicates) where the subject of a sentence is a non-nominative subject. In non-nominative subject constructions, the verb normally agrees with the theme/patient, if nominative case-marked. We will discuss how in the Munda languages the agreement patterns change when, in such cases, the subject is case-marked either locative, genitive, dative or accusative i.e. differently from what one would normally expect.

1.2 Agreement reversal

North Munda languages (Santali, Ho, and Mundari) have a rich agreement pattern where the predicate (verb or adjective) carries subject, direct and indirect object agreement markers. As we shall see, the subject agreement marker occurs either in the pre-verbal position or to the right of the indicative or [+finite] marker that every sentence carries. The object agreement marker occurs to the left of the [+finite] marker and its position is fixed, while the position of the subject agreement marker may vary depending upon the phenomenon of pro-drop in these languages, which is quite robust in these languages.

In this paper we wish to discuss a very unique phenomenon concerning agreement reversal in North Munda languages (Santali, Ho, and Mundari) in which the Subject Agreement Marker (SAM) in oblique object constructions occurs not in its own canonical position, rather it occurs in the position earmarked for object agreement, despite the fact that the predicate may be [-transitive]. We label this phenomenon as Agreement Reversal. Agreement Reversal in North Munda languages, though a robust phenomenon, has not drawn the attention of scholars thus far and, hence, remains unexplained. The aim of our paper is to explore the phenomenon of Agreement Reversal in Santali, Ho, and Mundari. We will argue that Agreement Reversal is the manifestation of NNS in these languages.

We argue that, with one exception, such reversal takes place not due to syntactic principles governing agreement alone, but due to thematic/ functional criteria because of which the nature of the predicate in a non-

nominative subject construction triggers such reversal. We hope to show that agreement, which is generally viewed in terms of syntactic principles governing constituent structure and analyzed in terms of hierarchical structures invoking the notion of c-command, may not be adequate to account for agreement reversal in the North Munda languages. In support of our hypothesis concerning the primacy of thematic/functional criteria, we present evidence concerning kinship constructions in which agreement reversal does not take place in one set of sentences. We'll show that when the possessor in kinship is case-marked by the genitive, agreement reversal takes place indicating that syntactic criteria too play a vital role. This, we argue, demonstrates that agreement in the North Munda languages can be accounted for by invoking syntactic as well as thematic/ functional criteria.

The common factor between NNS constructions and subject with genitive post-position is that in both cases the sentences involve agreement reversal. We observe that the occurrence of the incorporated genitive in the predicate does not correlate with the occurrence of the genitive in the subject position in Santali. Hence, at this point, it is appropriate to discuss the phenomenon of the incorporation of the genitive in such predicates. We label the incorporated genitive as the 'verbal genitive' to distinguish it from the genitive that occurs with a noun phrase, though both of them are homophonous. We further show that such verbal genitive incorporation is independent of the nature of case marking on the logical subject (an experiencer, possessor, recipient, or an 'undergoer' of an ailment or disease, etc.) We shall demonstrate that genitive incorporation in the predicate does not directly correlate with the occurrence of the genitive with a noun phrase and hence, they both are mutually independent of each other.

In certain specific thematic/functional domains mentioned in Section 2 below, the logical subject gets "demoted". We hypothesize that such demotion is made effective:

(i) Either by case-marking on the logical subject like in Indo-Aryan and Dravidian, and possibly, in languages like Icelandic and Russian. In Indo-Aryan and Dravidian languages, due to demotion, the case-marked logical subject does not control agreement.

Or

(ii) Even when it is nominative case-marked and *not* oblique case-marked like in Munda languages, the logical subject

still loses its prominence and gets demoted, not due to lexical case marking as such but due to functional/thematic criteria taking precedence over syntactic criteria, which is rather unique. The demotion is reflected syntactically in Munda languages by the Subject Agreement Marker (SAM) ending up as the Object Agreement Marker (OAM).

2. The Domains of Occurrence of the Non-Nominative Subject

South Asian languages, especially Indo-Aryan and Dravidian languages, are rich in constructions in which the subject is non-nominative case-marked. The functional/semantic domains in which a NNS occurs in South Asian languages are given below. The subject in these domains may be dative, accusative, locative, instrumental, or genitive.

Domains of occurrence of the non-nominative subject: (adapted from Subbarao & Bhaskararao 2004)

a. Psychological states and emotions
b. Physiological and mental ailments
c. Natural phenomena pertaining to body
d. Perceiver of visual and auditory actions
e. To express possession and kinship
f. Subject of predicates expressing obligation and necessity (desideratives)
g. To denote a recipient
h. Acquisition of knowledge or a skill
i. Part-whole relationship (single and double dative marking)

In the following section, we provide a brief description of verb agreement in Santali and agreement patterns in Ho and Mundari, which are quite similar to those in Santali.

3. Verb Agreement in Santali

The verb in Northern Munda languages exhibits agreement with the subject as well as the direct/indirect objects, subject to animacy conditions. In this paper, we show that in most of the domains mentioned above, the agreement pattern in Munda languages is *reversed*. The subject clitic, which is a *mobile* clitic, occurs in the *fixed* canonical position that is earmarked for a direct or indirect object. The logical subject in such cases

may, semantically, be an experiencer, a possessor, or an undergoer of an ailment or disease, etc. The subject in such cases is not an agent and the predicates that occur in such domains are all [-volitional]. In Agreement Reversal, the agreement clitic of the subject, whether it is nominative case-marked or non-nominative case-marked occupies the slot earmarked for the object in the functional domains mentioned above. We provide data from Santali, Mundari, and Ho to demonstrate this. Before we proceed further, a brief note on the agreement in Munda languages is a desideratum. We shall discuss the case of Santali agreement.

3.1 Subject and Direct Object agreement marking

In Santali, the subject, direct object, and indirect objects trigger agreement on the verb. A Subject Agreement Marker (SAM) is either suffixed to the *preverbal constituent* or suffixed to the *right of the finiteness marker* of the verb as in (1a) and (1b). Every declarative sentence carries the [+finite] marker or the indicative marker /*-a/* as the ultimate constituent, unless the SAM occurs to its right optionally, in which case the SAM becomes the penultimate constituent.

The Object Agreement Marker (OAM) /*-e-/* in the third person, for example, occurs to *the left* of the [+finite] sentence marker /*-a/* or on the preverbal constituent, the direct object *hopni* in (1a) and (1b). With [+transitive] verbs, the infix */-d-/* occurs to the right of the tense marker and to the left of the object agreement marker.

(1a) *iŋ* *hopni-**ŋ*** *ŋel-le-d-**e(y)-**a*
I Hopni-**SAM** see-pst-[+tr]-**OAM-**[+fin]
'I saw Hopni.'

(1b) *iŋ* *hopni* *ŋel-le-* *d-* ***e(y)-*** *a-* ***ŋi***
I Hopni see-pst-[+tr]-**OAM-**[+fin] -SAM
'I saw Hopni.'

It should be mentioned that while the position of the SAM is flexible and mobile, the position in which the OAM occurs is fixed and not mobile.

South Asian languages are pro-drop languages and, in an appropriate context, any argument or non-argument (adjunct) may be pro-dropped. Any argument or non-argument that occurs in a preverbal position will host the subject agreement clitic. In case the argument or non-argument is pro-dropped, the subject agreement clitic hops on to the constituent to its

left. So, when the direct object is pro-dropped, the subject agreement clitic occurs on the subject itself, as in (1c).

(1c) *iŋ$_i$-ŋ$_i$* *ŋel- le-* *d-* ***e(y)*** ***-a***
I-SAM see- pst-[+tr]-**OAM-**[+fin]
'I saw Hopni.'

If all arguments to the left of the verb are pro-dropped, it occurs to the right of the sentence marker *-a*. That is, in case the subject and object are both pro-dropped, the SAM occurs to the right of the [+finite] marker as in (1d).

(1d) *pro$_i$* *(pro)$_j$* *ŋel-le-d-**e(y)**$_j$-a-**ŋ**$_i$*
(I) (DO) see-pst-[+transitive]- **OAM-**[+fin]-SAM
'I saw Hopni.'

The [+finite] sentence marker which normally is the ultimate constituent in such cases becomes the penultimate constituent. Subbarao (2011) labels the process of such movement Clitic Hopping. The SAM may also occur to the right of the [+finite] marker, though no argument is pro-dropped as (1e) illustrates.

(1e) *bahā$_{(i)}$* *daka-ø* *jɔm-* *ke-* *d-* *a-* ***y**$_{(i)}$*
Baha rice eat- pst- [+tr] [+fin]-agr
'Baha ate rice.'

The *-ø* in (1e) indicates that the Subject Agreement Marker moved to the right of the final constituent, leaving this position empty.

3.2 Indirect Object Agreement Marking

The Indirect Object agreement clitic occurs in the same position as the direct object clitic does as (1f) and 1g) show. It thus occurs in the penultimate position to the left of the [+finite] sentence marker, unless the SAM occurs to the right of the [+finite] sentence marker.

(1f) *arɛl$_{(i)}$* *pɛŋa.**kɔ*** *$_{(j)}$-* *y$_{(i)}$* *ɔprɔm-* *kɛ- t-* ***kɔ**$_{(j)}$-* *a*
Arel guest.pl SAM introduce-pst- [+tr] OAM-[+fin]
'Arel introduced the guests to each other.' (field notes)

(1g) *iɲa* *ciṭhi-* *ɲa* *bheja* *-am-* *a*
I letter- SAM send -2 sg- [+fin]
'I'll send a letter to you.'
(Minegishi & Murmu 2001: 139-140)

bhej 'send' is a verb borrowed from Hindi-Urdu in Santali and is conjugated like an indigenous verb. In Santali, nouns borrowed from Hindi-Urdu too are conjugated like verbs, as there is no distinction between nouns, verbs, and adjectives in North Munda languages.[3]

(1h) *iɲa* *ciṭhi-* *ɲa* *bheja* ***am-*** *a*
I letter- SAM send **OAM**,2 sg (IO)- [+fin]
'I'll send a letter to you.'
(Minegishi & Murmu 2001: 139)
(iɲa 'I' and *iɲ* 'I' are alternants in Santali. cf. Minigeshi & Murmu 2001:141.)

4. The Role of Thematic/Semantic/Functional Criteria in Triggering Agreement

In the following discussion, we shall show that the functional/ semantic domains mentioned above play a crucial role in triggering agreement in North Muda languages (Santali, Mundari, and Ho). We've seen that in [+transitive] or [-transitive] sentences, the grammatical roles of the constituent play a very important role in triggering agreement. The positions in which the agreement marker occurs are fixed, and such occurrence of the agreement clitic is purely constrained by *syntactic criteria*. (cf. sentences 1a to 1h).

We shall now show that with regard to the lexical verb classes mentioned above in section 1, it is not syntax that plays the crucial role in triggering agreement, as the Subject Agreement Marker (SAM) does not occur in the canonical positions that are earmarked for it. Rather, it occurs in a position that is earmarked for the direct or indirect object, which is a *non-subject position*. Such paradoxical behavior of the SAM requires an explanation that we shall attempt to provide.

We first discuss the cases in which there is *agreement reversal or swapping* and will attempt to provide a plausible explanation to account for this seemingly contradictory behavior of the agreement clitic.

Subbarao (2012:177-178) argues that the predicate in sentences with a non-nominative subject in South Asian languages is *semantically bivalent* but *syntactically monovalent*. It is semantically *bivalent* because each

predicate requires an experiencer /possessor /recipient and a theme/ patient. The logical subject is non-nominative (locative, instrumental, dative, genitive, or accusative) case-marked. There is evidence to show that the predicate is [-transitive] and it cannot assign an accusative case to its theme patient.

The following examples in 4.1-4.2 from some select South Asian languages are illustrative:

4.1 Possession

Possession may be of two types: Alienable and inalienable. Body parts can be cited as a prime example of inalienable possession and the possession of concrete objects as a prime example of alienable possession. In sentences that express possession, the predicate is *be* in Dravidian, Indo-Aryan, Tibeto-Burman languages except for the Kuki-Chin languages. The subject in such cases is *not* nominative case-marked, but it is the non-nominative case-marked. In the inalienable possession example in (2), the subject is genitive case-marked, and the verb agrees with the theme which is nominative case-marked. The theme in (2) *pũũnch* 'tail', has the feminine grammatical gender.

Hindi-Urdu (Indo-Aryan)

(2)	*is*	*kutte*	*kī*	*lambī*	*pũũnch*	*thī*
	this	dogs (masc, pl)	gen	long	tail (fem.sg)	was (fem.sg)

'This dog had a long tail.'

In Telugu, the case marker that occurs with the subject depends upon whether the object of alienable possession is permanent or not. When it is permanent, the subject is dative case-marked as in (3).

Telugu (Dravidian)

(3)	*bahā*	*ki*	*cālā*	*dabbu*	*undi*
	Baha	dat	a lot of	money	be.pres

'Baha has a lot of money.'

When it is temporary, that is non-permanent, the subject is locative case-marked as in (4).

Telugu (Dravidian)

(4) *bahā daggira ī pustakālu lē.vu*
Baha near these books nm, pl not.nm, pl
'Baha does not have these books.'

In (3)-(4) above, the verb agrees *not* with the logical subject, namely the non-nominative case-marked possessor, but with the thing that is possessed, namely, the theme. The verb in Indo-Aryan and Dravidian languages agrees with the noun phrase that is nominative case-marked, and in the sentences above, the logical subject (possessor) is non-nominative case-marked. Hence, the predicate in such sentences agrees with the theme that is nominative case-marked.

4.2 Psychological (Psych-) predicates

Psych-predicates manifest mental states and emotions. The logical subject in such cases is an experiencer. In (5) below, the experiencer *Baha* is dative case-marked and hence does not trigger agreement. It is the theme *kōpālu.tāpālu* 'anger and the like' that triggers agreement, and it is in the non-masculine, plural. Hence, the verb exhibits non-masculine[4], plural agreement with the theme.

Telugu (Dravidian)

(5) *bahā ki kōpālu.tāpālu ekkuva.gā unṭā.yi*
Baha dat anger and the like a lot be.pres.nm.pl
'Baha has a lot of anger and the like.'

A similar agreement pattern is manifested in other domains too. (See Subbarao 2012 for further details).

In Tibeto-Burman languages, it is only in the Kuki-Chin languages that the verb agrees not only with the subject but also with the direct object as well as the indirect object. In all the semantic domains mentioned above, it is the logical subject whether it is nominative case-marked or non-nominative case-marked that controls agreement in Tibeto-Burman languages.

To sum up the discussion above, it is the nominative case-marked NP that controls and triggers agreement in Indo-Aryan and Dravidian languages, and the logical subject (an experiencer or possessor, for example) that is non-nominative case-marked does not control and trigger agreement. That is, the agreement is a purely structural and grammatical phenomenon in Indo-Aryan and Dravidian languages. With this background in mind, let us now consider instances of possession in Santali. We wish to demonstrate that Santali, Mundari, and Ho do not adhere to the structural principles of the agreement discussed above with regard to the domains mentioned above. We'll show that it is the functional/thematic domains that are crucial for agreement reversal in Munda languages.

5. Possession in Santali

5.1 Possession of Concrete Objects in Santhali

As mentioned earlier, Santali has a subject, direct object, and indirect object agreement. The verb *mena?* ~ *mena* 'be' indicates possession too, just as the verb *be* in many South Asian languages does. In fact, except for the Tibeto-Burman languages of the Kuki-Chin[5] family, it is the verb *be* that manifests possession in all other South Asian languages. Santali exhibits two sets of patterns with regard to possession. We shall label them as agreeing pattern with *mena?* - 'be' and non-agreeing pattern with *mena?* - 'be'.

5.2 Agreeing Pattern

In the agreeing pattern with *mena?* -'be', the subject is genitive case-marked, and the verb too carries the incorporated genitive postposition *-ta-*,that occurs to the right of the verb stem *mena?* -'be'. Though the subject is genitive case-marked, the verb exhibits agreement with the subject. This shows that lexical case marking on the subject has no bearing on an agreement on the verb.[6] Further, the subject agreement marker (SAM) *iɲ-* '1sg' (in bold in (6)) occurs in a position that is earmarked for the *object* i.e., to the left of the [+fin] marker. Though it occurs in the object position, it is not a subcategorized argument of the verb *mena?* ~ *mena* -'be', as it is [-transitive] and a [-transitive] predicate cannot permit an object to occur. Thus, such occurrence leads to a paradoxical situation, where the possessor ends up as an object of a [-transitive] verb.

Agreeing pattern with *menaʔ* -'be'-genitive incorporated

(6) *iɲ-ak'* *kitəb* *menaʔ-* ***t-*** ***iɲ-*** *a*
I-gen book be- gen- **1sg-** [+fin]
'I've a book.'
(Minegishi & Murmu 2001: 151)

In (7) below, though, the subject is pro-dropped, the subject agreement clitic *–am-* '2 sg' occurs to the *right* of the incorporated genitive in the canonical position earmarked for the object.

Genitive incorporated-subject pro-dropped[7]:

(7) *tiɲ-ək'* *kitab* *menaʔ-* ***t*** ***am-*** *a*
How many book be- gen- 2 sg- [+fin]
'How many books do you have?'

Here is another example where the predicate exhibits agreement with the subject.

Subject Genitive Case-marked:

(8) *iɲ-ak'* *bariya* *puthi* *menaʔ-* ***t-*** ***iɲ-*** *a*
I-gen two book be- gen- 1sg- [+fin]
'I have two books.'(field work)

5.3 Non-Agreeing Pattern

We shall now discuss some instances where the verb does *not* exhibit subject agreement and neither is there genitive incorporation in the verb. In contrast to (8) above, the incorporated genitive *-t-* and the 1st person subject agreement marker clitic */-ɲ/* are absent in (9) below.

Non- Agreeing Pattern – no genitive incorporation in the verb:

(9) *iɲ-ak'* *kitəb* *menaʔ-* *a*
I-gen book be- [+fin]
'I've a book.'
(Minegishi & Murmu 2001: 150)

The forms of the verb *be, exist* in Santali are *mena* and *menaʔ.* These are phonologically conditioned allomorphs. The verb stem is *mena*. When a consonant follows *mena* 'be', the glottal stop occurs to the right of *mena* 'be' yielding *menaʔ* 'be'. The fact that in (9) and (10) below, *mena* 'be' has the form *menaʔ* 'be' with a glottal stop indicates that the incorporated genitive and the SAM are dropped.

(10) *iɲ-ak'* *bāryā* *kitəb* *menaʔ-* *a*
I-gen two book be- [+fin]
'I've two books.'
(Minegishi & Murmu 2001: 150)

That is, the Agreeing Pattern is the normal pattern and the Non-Agreeing Pattern is a derived pattern. The glottal stop on *mena,* we hypothesize, is a vestige left over from the Agreeing Pattern.

In the following section, we shall consider data where kinship relations are manifested.

5.4 Possession: Kinship

The expression of kinship relations plays a very crucial role in South Asian languages. In Hindi-Urdu and Punjabi, for example, the subject is genitive case-marked when a kinship relation is expressed indicating that expressing kinship through language is an exclusively significant notion.

In Santali too, kinship has a critical role to play, as the expected agreement reversal does not take place when a kinship relation is manifested.

In (11) below the logical subject *iɲ* 'I' is the genitive case-marked in Santali and interestingly, the genitive *ta* is not incorporated in the verb, which shows that the genitive may be *optional* when the patient manifests a kinship relation. In this, there is ***no agreement reversal***, as one would expect. The expected pattern is for the possessor *iɲ* 'I' to manifest agreement in the position earmarked for the object in (11). However, it is not the case as (11), and (12), below demonstrate.

In (11a)-(12a), the agreement marker for the possessed noun phrase occurs in its canonical position, and there is no Subject Agreement Marker. (11b)-(12b) is ungrammatical as the possessor occurs in the canonical position of the object.

(11a) *iɲ- rin$_i$ bārya boyha$_j$ menaʔ- kin$_j$ - wa
I- gen two brother be- dual- [+fin]
'I've two brothers.' (Minegishi & Murmu 2001: 151)

(11b) *iɲ- rin$_i$ bārya boyha$_j$ menaʔ- iɲ$_i$ - wa
I- gen two brother be- 1 sg- [+fin]

(12a) ām- rin$_i$ peya boyha$_j$ menāʔ- ko$_j$ - wa
you- gen three brother be- 3pl- [+fin]
'You've three brothers.' (Minegishi & Murmu 2001: 152)

(12b) *ām- rin$_i$ peya boyha$_j$ menāʔ- am$_i$ - wa
you- gen three brother be- 2sg- [+fin]

Sentences (11a) and (12a) show that in the expression of kinship relations, the subject does not lose its primacy either in terms of agreement reversal or case marking on the logical subject.

In sentences (13) and (14), there is no agreement marker at all. Such absence of the agreement marker is rather unexpected, and it needs to be further explored.

(13) iɲ- ic' mid' boyha mena- ya
I- gen two brother be- [+fin]
'I've two brothers.' (Minegishi & Murmu 2001: 151)

(14) uni- **rɛn** moṛɛ~gɔtāŋ hɔpɔnɛra mɛnaʔ- **tā-** yā
she- **gen** five.Cl daughters be- **gen-** [+fin]
'She has five daughters.' (Field notes)

Example (15) is a case of a non-nominative subject construction.

Possession: Kinship – Agreement Reversal:

(15) əliɲ- ren- dɔ mɔṛɛ~- gɔṭɛn hɔpɔn menak'-
we (dual,excl)- gen- top five- cl son exist-
ko- ta- liɲ- a
3pl (OAM)- gen- 3 dual (SAM)- [+fin]
'We've five sons.' (Folktales 3:26) (From Neukom 2001:169)

The following points are worth mentioning regarding (15).

(i) There is agreement reversal in (15).
(ii) The logical subject, that is, the possessor is genitive case-marked.
(iii) The SAM *liŋ* of the possessor occurs in the canonical position of the object to the left of the [+finite] marker.
(iv) The OAM *ko* occurs to the right of the verb *menak'* 'be'.

5.4.1 (No) Agreement Reversal – Verbal genitive *–ta-* present

There are some instances in which, despite the occurrence of the verbal genitive *-ta-*, agreement reversal does not necessarily take place.

No Agreement Reversal- Verbal genitive *-ta-* present:

(16) *thɔɽɔ-* *rɛn* *gidrəᵢ-* *menaʔ-* *kɔᵢ-* *tā-* *(y)a*
Thoro- gen,pl child- has- pl- [+tr]- [+fin]
'Thoro has children.' (field notes)

In contrast, in (17) and (18) agreement reversal takes place, and the subject is pro-dropped.

Agreement Reversal - Verbal genitive *–ta-* present:

(17) *gidrə* *menak'-* *ko-* *ti-* *ŋ-* *a*
child be- 3pl- of, gen- 1sg- [+fin]
'I have children.' (field notes)

(18) *girdrə-* *(kɔ)* *menaʔ-* *ko-* *t(a)-* *ŋ-* *a*
child- pl have- 3 pl- gen- 1sg- [+fin]
'I've children.'

To sum up the discussion above, it appears that the occurrence or non-occurrence of the genitive has no direct correlation with Agreement Reversal. We propose to account for the phenomenon of Agreement Reversal and its absence by considering more data.

5.4.2 Possession: Abstract

In Santhali, sentences in which the possession of a quality is expressed, and the predicate is *mena* 'exist, be', the possessor of the quality is in the *locative cas*e and the predicate carries the incorporated genitive *-ta-*.

5.4.3 Psychological (Psych-) predicates manifesting psychological states

Note that the verbal genitive *-ta-* is incorporated in (19), though the subject is locative case-marked. This shows that the occurrence of the verbal genitive *-ta-* is independent of the case marking on the subject.

(19)	*uni-*	*re*	*endre*	*mena-*	*k'-*	*ta-*
	he-	*loc*	anger	exist-	3 [-animate]-	*gen-*
	e-	a				
	3 sg. poss-	[+fin]				

'He is angry with him (or: he has anger on him.)'
(Santali Dictionary II.293), (as quoted in Neukom 2001)

We've seen above verbal genitive *-ta-* occurs:

(i) When the subject is nominative case- marked as in (17),
(ii) When the subject is genitive case- marked as in (16), and also,
(iii) When the subject is locative case- marked as in (19).

This leads us to the conclusion that the occurrence of the verbal genitive in sentences expressing kinship and psychological states is not dependent on the case-marking on the logical subject, which is either a possessor or an experiencer.

6. Agreement Reversal in the Psychological States and Emotions

There is no distinction between nouns, verbs, and adjectives in North Munda languages. A psych-expression such as *khušī* 'happiness', a noun borrowed from Hindi-Urdu in Santali is conjugated like a verb. There is a semantic shift in the predicate *khušī* 'happiness' and it has the interpretation of 'like'. The sibilant /š/ of Hindi-Urdu changes to /s/, a common pattern found in many South Asian languages.

Note that *khusi* 'like' which is semantically bivalent takes the middle [-transitive] marker when used as a verb in Santali. According to our analysis, the verb *khusi* 'like' comes under the domain of 'psychological states and emotions'. Interestingly, in Indo-Aryan and Dravidian languages, such predicates are [-transitive] and they invariably take either a dative in all Dravidian languages and in languages such as Hindi-Urdu, Punjabi, and Kashmiri or a genitive in Bangla, Assamese, etc. In Santali

too, the predicate is treated as an intransitive predicate. There is no agreement reversal with such predicates. The SAM *-m* (in bold in (20)) occurs in the pre-verbal position on the topic marker *dɔ*. If it were a case of agreement reversal, as the SAM *-m* would occur to the right of the [-transitive] marker *-k'*.

Santali

(20) *nui- dɔ- **m** khusi- k'- khan*
this- top- 2sg like- [-tr]- if
'If you like it - - -' (Folktales 9:118) (From Neukom 2001:39)

6.1 Auditory and Visual Perceptions

In sentences with predicates expressing the notion of audible or visible, the subject in Dravidian and Indo-Aryan languages is in the non-nominative case and the predicate is [-transitive] and non-volitional.

The predicate *ɲel. oʔka* 'appear' in Santali comes under the category of visual perceptions. Just as in Dravidian and Indo-Aryan languages in Santali too, the predicate is [-transitive], and hence, the [-transitive] marker */-oʔk* / (in bold below) occurs in the predicate. Note that there is no agreement reversal in (21).

(21) *uni kuṛi əṛi mõj ɲel.**oʔk** kān-* a
That girl very beautiful see[-tr] pres- [+fin]
'The girl is looking/[appears] very charming.'
(Minegishi & Murmu, 2001: 154)

6.2 Ailments

When an ailment is expressed, and the predicate manifests a process, the predicate *ɲām* 'get' is used and the subject is in the nominative case. In such cases, there is no genitive incorporation as (22) below shows. The significant point to be noted is that though the subject is in the nominative case and is the initial argument of the sentence, the agreement clitic (in bold) occurs in the position earmarked for the direct/indirect object.

Physical ailments (process)–no genitive incorporated:

(22) *baha$_i$ ruəʔ ɲam^{9}- aka- d- **e$_i$**- y- a*
Baha fever get- pst- [tr]- **3 sg**- glide- [+fin]
'Baha caught a fever.' (Field notes)

(23) *gidrə- (kɔ)i ruek' ɲam- aka-t'- **ko** i- w- a*
Child- pl fever get- pst-[-tr]- **[+pl]**- glide- [+fin]
'The children caught fever.'

(24) *pɛ pon māhā-ge bɔhɔk'haso-ɲɔːk'-le- d- **e**- a*
three four day-foc head ache-little -pluperf-[+tr]-SAM-[+fin]
'She suffered a little from a headache three or four days.'
(Bodding Folk Tales 21:274)

Having discussed the cases in which agreement reversal takes place, or does not, we shall attempt to show that subject (experiencer/possessor), still possesses all the properties attributed to a subject, despite being relegated to a non-subject position with regard to an agreement due to the phenomenon of Agreement Reversal.

6.3 Subject Properties of a Non-Nominative Subject in Munda Languages

(i) In Munda languages, in all the domains mentioned above, the subject is the first c-commanding NP in a sentence.

(ii) In sentences with psychological predicates, the Subject is in the nominative case as in (25) below and triggers agreement. It is the antecedent to anaphoric elements like *at'/aprē*. The intransitive marker *-n-* occurs with the verb.

(25) *upəl$_i$ at'/ãprē$_i$-y icetānrē bejār-aka-n-a*
Upəl self-SAM on angry-pst-[-tr]-[+fin]
'Upel was angry/ upset with herself.'

(iii) PRO, the notional subject of the conjunctive participial clause invariably requires the subject as its controller in the matrix clause. We provide (26) as an illustration to show that the subject is the controller of PRO.

(26) [*un.kin$_i$ [PRO$_i$ rābāŋ ɲām-akā- t- kin i- te]*
They (dual) cold get- perf pple -[+tr]- dual (SAM)-cpm
orāk sen-en-ā]
house go-pst-[+fin]
'Having caught a cold, they (dual) went home.'

The issues that need an explanation in (26) are the following: *ɲam* 'have, get' is a [+transitive] verb, and consequently, the [+transitive] marker *-t/d-*occurs with the verb. What is crucial to observe is that the notional subject of the embedded clause is PRO, which according to standard assumptions, is *uncase-marked and ungoverned*. PRO, despite being the notional subject and uncase-marked, triggers agreement on the verb. Here again, the expected agreement pattern is for the subject of the [+transitive] verb to occur either in the preverbal position or to the right of the [+finite] marker at the end of the sentence. This is contrary to the expected pattern, as the embedded predicate *rābāŋ ɲām* 'cold get' manifests a physical ailment. Hence, the SAM occurs in the position earmarked for the object due to the phenomenon of agreement reversal.

7. Some Issues

There is an example where the dual agreement marker *kin* alternates with the plural agreement marker *ko,* which we did not find during our field work, a counter-example to the normal pattern in Santali:

(27)	*iɲ-*	*rin*	*barya*	*boyha*	*menaʔ-*	***kin/ko-***	*a*
	I-	gen	two	brothers	be (loc)-	**dual/pl-**	[+fin]

'I've two brothers.'
(Minegishi & Murmu 2001:151)

We provide below some examples from Ho and Mundari to show that Agreement Reversal takes place in these languages too.

8. Agreement Reversal: The Case of Ho

In Ho too agreement reversal/swapping takes place just as it does in Santali. The following quote from Burrows (1980:86) is illustrative. Burrows labels this as an impersonal construction. "When conjugated impersonally, the pronominal sign [agreement-SDKE] denoting the person who experiences the physical or mental condition is inserted in the verb in the same way as the *animate object-sign* [emphasis provided-SDKE] of transitive verbs." (Burrows, 1980:86).

In (28) below, the Subject Agreement Marker *-iŋ-* occurs to the *right* of the predicate *suku* 'please [-transitive]' in a position that is earmarked for the direct/indirect object.

"Impersonal construction":

(28) *pro* *suku-* ***iŋ-*** *tan-* *a'*
(it) please- **1sg-** pres- [+fin]
'It pleases me.' (Burrows, ibid)

There is a corresponding structure, according to Burrows, which he labels as personal construction. In (29) below, the Subject Agreement Marker *-iŋ-* occurs to the **right** of the [+finite] sentence marker *–a*, as the pre-verbal constituent is pro, which is null.

"Personal construction":

(29) *pro* *suku-* *tan-* *a'-* ***iŋ***
please- pres- [+fin]- **1 sg**
'I'm pleased.' (Burrows ibid)
(Literally: 'Something pleased me')

In (30) below, the 2nd person agreement marker *me* occurs to the **left** of the [+finite] sentence marker *–a* in the object position.

(30) *giu-* *ke-* *d-* ***me-*** *a*
shame (verb)- pst- [+tr]- **2 sg-** [+fin]
'It shamed you' or 'you were ashamed.'

9. Agreement Reversal: The Case of Mundari

Mundari too exhibits Agreement Reversal, just as Santali and Ho do. In (31) below, the predicate *manda nam* 'to catch a cold' manifests a physical ailment and it comes under the domains that we've mentioned above. The subject agreement clitic (SAM) occurs in the position of the object agreement clitic in (31) to the left of the [+finite] marker *–a.*

(31) *lum-* *ja-* *n-* *ci* *manda* *nam-*
get- wet- [tr]- cpm cold- get-
ja- *ʔ-* ***ɲ-*** *a*
perf- pst- **1 sg-** [+fin]
'As I got wet, I got a cold.' (Osada 1992:108)

A similar phenomenon is manifested in the following sentences. We've marked the SAM occurring in the OAM position in bold.

(32) *duṛum* *ki-* *ʔ-* ***ɲ-*** *a*
sleep- perf- pst- **1 sg-** [+fin]
'I felt sleepy.' (Osada 1992:108)

(33) *reŋgʔ-* *ja-* *ʔ-* ***ɲ-*** *a* *ja-*
hungry- perf- pst- **1 sg-** [+fin] perf-
'I am hungry.' (Osada 1992:109)

A predicate such as *happy,* like many such predicates which express emotions may be used with a [+volitional] interpretation or with a [-volitional] interpretation in South Asian languages. When expressed in the [+volitional] sense, the subject in South Asian languages is in the nominative case and when expressed in the [-volitional] sense, it is either in dative or genitive case in Dravidian and Indo-Aryan languages. It appears that Mundari also makes this distinction in terms of volitionality. The data is from Osada (1992: 106). The following sentences form a good syntactic minimal pair.

Example (34) below is a sentence in which the predicate is [+volitional], and hence, it follows the normal agreement pattern where the SAM occurs to the right of the [+finite] marker.

(34) *pro* *suku-* *le-* *n-* *a-* ***ko***
(they) happy- pst- [-tr]- [+fin]- **3 pl**
'They had been happy.' (Osada 1992: 106)

Example (35) below is a sentence in which the subject is an experiencer, and hence, the predicate is [-volitional]. With a non-volitional predicate, there is agreement reversal and the 3rd person plural agreement marker occurs to the **left** of the [+finite] marker.

(35) *pro* *suku-* *le-* *d-* *ko-* *a*
(they) happy- pst- [+tr?] **3 pl-** [+fin]
'They had been happy.' (Osada 1992:106)

For the above minimal pair, the translation provided by Osada (ibid) however, does not reflect this distinction, and hence, the interpretations

need to be verified. The fact remains that there is Agreement Reversal in Mundari too.

10. Conclusion

We've discussed a very unique phenomenon concerning agreement reversal in North Munda languages (Santali, Ho, and Mundari) in which the Subject Agreement Marker (SAM) in oblique object constructions occurs not in its own canonical position but in the position earmarked for an object, despite the fact, the predicate may be [-transitive]. We labelled this phenomenon as Agreement Reversal. We argued that such reversal takes place not due to syntactic principles governing the agreement, but due to thematic/functional criteria due to which the nature of the predicate in a NNS construction triggers such reversal. We have shown that agreement which is generally viewed in terms of syntactic principles governing constituent structure and is analyzed purely in terms of hierarchical structures invoking the notion of c-command may not be adequate to account for agreement reversal in the North Munda languages. In support of our hypothesis concerning the primacy of thematic/functional criteria, we presented evidence from data concerning kinship constructions in which agreement reversal does not take place in one set of sentences. We've shown that when the possessor in kinship is case-marked by the genitive, agreement reversal takes place indicating that syntactic criteria too play a vital role. This, we argue, demonstrates that agreement in the North Munda languages can be accounted for by invoking syntactic as well as thematic/functional criteria.

Abbreviations

[+fin]: finite
[+pl]: plural
[+tr]: transitive
[-fin]: non-finite
[-tr]: intransitive
1: first person
2: second person
3: third person
cl: classifier
cpm: conjunctive participle marker
dat: dative
dual: dual
excl: exclusive
fem: feminine
foc: focus
gen: genitive
IO: Indirect Object
masc: masculine
nm: non-masculine
OAM: Object Agreement Marker
perf pple: perfect participle
perf: perfect

pluperf: pluperfect
pst: past
sg: singular
pres: present
SAM: Subject Agreement Marker
top: topic

Notes

* The fieldwork data were collected by the first author and Mayuri Dilip in Shantiniketan, West Bengal and Ranchi, Jharkhand. The funding was provided by a research grant from Utrecht University.

[1] Some languages of the Naga subgroup of Tibeto-Burman languages too have the NNS construction.

[2] There are four languages that we know of where the theme/patient may also be the accusative case-marked. These include Bangla, Assamese, Bodo, Tamil, and Malayalam. (See Subbarao 2012: 171-178).

[3] See the statement of Deeney (1979: iiiv) about Ho, which is equally applicable to Santali and Mundari. Kharia too does not exhibit any such difference between nouns, verbs, and adjectives. See Peterson (2013) for Kharia.

[4] The gender distinction in Telugu is between masculine and non-masculine. Feminine and neuter nouns are placed in the category of non-masculine.

[5] Kuki-Chin languages have three distinct verbs for *be (equational), be (locational),* and *have*.

[6] In most of the Indo-Aryan languages and in all the Dravidian languages a case-marked NP cannot trigger agreement. See Subbarao 2012 for further details.

[7] Nukom (2001:47) glosses *-ək*' as nominalizer.

[8] *ɲam* 'get' also means 'meet' and *ɲa-pa-m* 'meet$_1$-vrec-meet$_2$' is the reciprocal form of the verb.

References

Bhaskararao, P. and K. V. Subbarao, eds. *Non-nominative Subjects, vols. I and II.* Amsterdam/Philadelphia: John Benjamins, 2004.

Bodding, P.O. *A Santali Grammar for Beginners*. Benagaria, Santal Mission of the Northern Churches, 1952.

—, *Santali Folk Tales*. Oslo: H. Aschehogue & Co, 1925-29.
(Data taken from Neukom,2001).

Burrows, L. *The Grammar of the Ho Language: An Eastern Himalayan Dialect* New Delhi: Cosmos Publications, 1915/1980.

Butt, M., and A. Deo, "A Historical Perspective on Dative Subjects in Indo-Aryan." Paper presented at the LFG 13, Hungary, University of Debrecen, July 2013.

Deeney, J. *Ho Grammar and Vocabulary,* Chaibasa: Xavier Ho publications, 1979.

Donegan, P. and D. Stampe. “Rhythm and the Synthetic Drift of Munda”. In *The Yearbook of South Asian Languages*, edited by Rajendra Singh,3-36, New Delhi: Sage Publication, 2004.

Macphail, R. M. *An Introduction to Santali*. Calcutta: Firma KLM Private Limited, 1964.

Neukom, L. *Santali*. Lincom Europa, 2001.

Peterson John. *A Grammar of Kharia*. Leiden: Brill, 2011.

—. “Parts of Speech in Kharia. A formal account.” In *Flexible Word Class: Typological studies of underspecified parts of speech,* edited by J. Rijkhoff and E. van Lier,131-168. Oxford: Oxford University Press, 2013.

Mohanan, K. P. and T. Mohanan. “Dative Subjects in Malayalam: Semantic Information in Syntax.” In *Experiencer Subjects in South Asian Languages* edited by M. K. Verma and K.P. Mohanan, 43-58, Stanford: CSLI Publication, 1990.

Osada, T., *A Reference Grammar of Mundari*, ILCAA, Tokyo: Tokyo University of Foreign Studies, 1992.

Pandharipande, R. “Experiencer (dative) NPs in Marathi.” In *Experiencer Subjects in South Asian Languages* edited by M. K. Verma and K.P. Mohanan, 161-79, Stanford: CSLI Publication, 1990.

Subbarao, K.V. “Aspects of Santali (Munda) Grammatical Structure.” Paper presented at the *Seminar on Munda Languages*. Hyderabad, University of Hyderabad, 2011.

—. *South Asian languages: A syntactic typology*. Cambridge: New York, Delhi: Cambridge University Press, 2012. *Web material (soft copy).* Free download available@ www.Cambridge.org/subbarao>. Cambridge: Cambridge University Press.

Verma, M.K. and K.P. Mohanan, eds. *Experiencer Subjects in South Asian Languages.* Stanford, CSLI Publications, 1990.

CHAPTER FOUR

PERFORMANCE IN ELICITATION: METHODOLOGICAL CONSIDERATIONS IN THE STUDY OF MUNDARI EXPRESSIVES

NATHAN BADENOCH, NISHAANT CHOKSI, MADHU PURTI, AND TOSHIKI OSADA

1. Introduction

Mundari expressives are marked, distinct from other words through reduplication not motivated by semantics or iconicity (Osada 2010; Phillips and Harrison, 2017). This markedness allows expressives to stand out in the speech stream, endowing them with heightened performativity (Lahti, Barrett, Webster 2014: 335). As expressives are not referential, we view them as a "depictive" mode of signification: "they show rather than tell, depict rather than describe, enact rather than simply refer" (Dingemanse 2014, 387), and they involve "multiple semiotic resources including speech, manual gestures, and other forms of visible bodily behavior" (Dingemanse 2014, 385). Since expressives involve embodied performance and their meanings are context-situated, they are very difficult to elicit (Diffloth 1972).In this study, we worked with Mundari-speaking consultant Madhu Purti and co-author in Kyoto, Japan on two separate tasks. In the first task, we video-recorded her telling stories, which we then transcribed analyzing how the expressives were deployed within the narrative and as part of creative performance. In the second task, we took a list of Mundari expressives and discussed the meanings with Purti in a conversation. This task was also video-recorded so we could assess what gestures and sounds coordinated with the use of a particular expressive. In the process of transcribing, analyzing, and discussing the expressives, we noticed several areas of markedness that run through her expressive performance. We begin by providing some

background into expressive morphology and syntax and then we provide some examples of our performance elicitation to describe how grammar combines with performance elements such as gesture, echo, phonation, etc. to provide an overall picture of expressive markedness in Mundari.

2. Mundari Expressives

Mundari has a rich system of expressives which display both syntactic flexibility and morphological specificity. Morphologically, they are marked through processes of reduplication, which includes identical reduplication, partial reduplication, and vowel mutation (Osada 2010; Osada, Purti and Badenoch 2020, Badenoch and Osada 2020).

Examples of identical reduplication include:

cakob-cakob	eating noisily
lugum-lugum	mumbling or chewing
gusu-gusu	an inactive character
suyuŋ-suyuŋ	lean and small body (person)
mugui?-mugui?	smiling cheerfully

Identical reduplication has a similar morphology to the phenomenon of "echo words" in Indian languages (e.g. Mohan 2006). Yet unlike echo-words, expressives in Mundari can also be subject to consonant and vowel mutations in the reduplicative morphology. In grammatical terms, this switch heightens the markedness of the words. Examples of consonant mutation in expressives include:

adil-padil	'spacious, not densely occupied; said of the space in a house, where there is ample room to sit or move around'
caɽa-paɽa	'sound of fire popping as it burns'
hejo-bejo	'unable to speak properly, unclear pronunciation of a child'
kuca-muca	'long and twisting, like a road or river'
rawa-dawa	'doing things without concern for others'

Examples of vowel mutation include:

baʔ-buʔ	‘exposed to fire, resulting in a hole being burned in the surface, refers to cloth or item of clothing'
carbaɻaʔ-curbuɻuʔ	‘repeated sound of hand splashing into water and sinking into mud, when planting rice '
kabaʔ-kobоʔ	‘walking slowly and stiffly, slightly disoriented or uncomfortable, such as after sleep or illness '
lagar-logor	‘speaking while crying, describes the shape and wetness of mouth;
palad-pilid	‘sudden flashes of light, repeated and punctuated'
sikae-sokoe	‘loose and moving slightly in place with a rattle’

Mundari expressives also display markedness with respect to syntax. Expressives can occupy any place in the sentence, i.e., in a predicate, complement, or argument slot. As the head of the predicate, expressives can take derivational suffixes such as passive. Expressives usually connect with some notions, actions, states, and internal and external senses. If it connects with actions, it occupies the complement slot with *-ta-n* (Aspect Marker + Indirect Marker). It can modify a verb or form a serial verb such as in:

busuʔ-re	*seta-hon=e*	***utul-putul-ta-n-a***
straw-LOC	dog-child-3Sg-S	EXPR-PROG-INTR-IND

“The puppy is rolling in the straw, disappearing and reappearing.”

A few expressives can occupy the argument slot to modify a noun or noun phrase.

iniʔ-aʔ	***isiɻi-sikiɻi***	*ka-m*	*suku-a*
that person-GEN	EXP	NEG-2Sg	like-IND.

“You don’t like that person who always laughs to herself suspiciously.”

The following sections focus on data recorded in our recording and elicitation sessions with Purti.

3. Performance in Elicitation

3.1 Storytelling

We recorded four narratives with Mundari consultant Madhu Purti, and co-authors in Kyoto and found that expressives were used to highlight the actions of characters within the narrative. These were often accompanied by parallelism, phonation, speed change, and gesture. We discovered the following features that contribute to how expressives are deployed depictively in Mundari and could be explored in greater depth in further research.

3.2 Expressive Echo

Munda expressives are based on reduplicated pair parts, with certain consonants "echoing" both within the word and, when deployed in a narrative, within the phrase structure as well (Osada, Purti and Badenoch 2020). This suggests that Mundari may have an aesthetic principle of symmetry or parallelism that has been shown to be present in Austro-Asiatic languages like Khmer (Haiman 2011).

ekdamtuyu boro-ja-d-a= eʔ, naaʔdo[ena-te]boro-ja-d-ci cika-yaɽoʔ
ena-te jajuɳu ja muli-te-ge tuyu ***raʈa-paʈa****-ta-n=eidi-ja-d-a*
*bir bir hana****aʈa-maʈa*** *bir-ko tala-re ena-te*

"Suddenly the jackal – the boy became scared now even though he was scared, [he was thinking] what will he do, from somewhere the bush [he was on] moving straight, crackling [**raʈa-paʈa**] as the tiger carried him through the thick [**aʈa-maʈa**] jungle."

In this utterance taken from a larger narrative, we find retroflex /ʈ/ echoed throughout the utterance, in the pair-parts *raʈa-paʈa* and *aʈa-maʈa*, and by the adjacent placement of these expressives in the clause. Though the meanings of the expressives are different, as well as the senses to which they refer (crackling noise vs thick, dark), equivalences are made through sound that carries through the utterance. The meaning, therefore, is built up, made through reverberation of sound as well as a contrast of sense and depicted through phonology, morphology, and syntax.

The /-aʈa/ sequence provides a rhyming pattern that links the key expressive foci of the passage. The crackling sound *raʈa-paʈa* of the movement is a commonly heard expressive, which evokes not only the sound but also small movements of hands or paws that create the sound.

The second expressive element *aʈa-maʈa* depicting the dark depth of the forest is found only in the storytelling register. Thus, while the rhyme of the expressive sequence gives a desired poetic effect through sound, the parallelism of the frequent-rare expressive pair provides a cultural reference to link the directly imagined character actions with a more mythical natural setting.

3.3 Expressive Gesture

Expressives are also routinely accompanied by a gesture. This has been increasingly recognized in a recent study of expressives. As such, it is a common form of multimodal performance. The following utterance is taken from a narrative discussing the story of a boy and a jackal:

kosa-ate=m deʔ ke-n-a met-a-i-ta-n-a
landi-saʔ-te=ɲ deʔ-ke-n-a met-a-i-ta-n-a
ena-te ***rago-pago rago-pago rago-pago rago-pago*** *ja-d-a -eʔ*
oɽoʔ ka-e deʔ-daɽi-e-a

"From which side did you climb, he [the tiger] is asking him. From the back, he tells him, so the tiger tries to claw his way up but he cannot climb up."

In this utterance, the expressive *rago-pago* is echoed four times, creating an iterative effect both within the pair-part as well as in the clause. The expressive is also glossed by the speaker (metapragmatically) through the use of the climbing action depicted with the motion of her arms and head. The repeated use of climbing adds a level of depiction directing the audience in how to interpret the meaning, while also allowing the narrator to inhabit the role of the character. The long repetition of *rago-pago* emphasizes the effort of the tiger, a common form of iterative marking in expressives, which was in the end, futile. Audience attention is drawn to the utterance through repetition and gestural glossing.

3.4 Phonation

Expressives in the narrative were also found to be marked by heightened phonation. Breathy phonation is a common part of the phonological analysis in Austroasiatic languages although the feature may be disregarded in an analysis of narrative and pragmatics. In Munda languages, breathy phonation is not considered to be a salient phonemic

characteristic. In our data, however, the use of breathy voice appears to be an articulatory tool used to foreground the expressive meaning.

ena-te do
oɽoʔ ragosa ***dạ̤ɽạ̤-du̱ɽu̱ dạ̤ɽạ̤-du̱ɽu̱****-ta-n=e nir-au-ja-d-a*
diri-ko ekdam guli-yoʔ leka ge

"So, then the demon returned home, running very fast [**dạ̤ɽa-du̱ɽu dạ̤ɽa-du̱ɽu**] and [smashing into] the stones, as if he was stumbling".

While this utterance begins at slow speed, the speech speed quickens around the time of the expressive, giving a vocal depiction of fast speed which the expressive implies. The retroflex flap also facilitates faster speech, while vowel alternation calls attention to the action. There is no phonation at the (prosaic) beginning of the utterance while phonation is heard during the expressive and carried through to the subsequent 'stones' /diri/ part of the utterance. While the speed is increased to evoke imagery of fast running, the breathiness seems to serve to enhance a certain heaviness of the running, providing a subtle accent to the demon's presence, while at the same time perhaps transferring part of the expressiveness from the running onto the movement of the stones through the extension of breathiness. It is possible that the breathiness is motivated phonetically by the voiced /d/, but the performance of the utterance foregrounds the narrator's excitement.

Pitch changes accompanying expressive use have been discussed by Nuckols for Quechua, signalling the importance of secondary articulatory characteristics of this mode of speech (Nuckolls 1996). This raises the question of phonation in Mundari grammar or at least within the expressive syntax of the language.

3.5 Expressive Serialization

During the course of the narration, expressives were found to occur in a serial sequence. For example:

ena-te do oɽaʔ-te uruŋ-ja-n ci
duɽa oʈaŋ-ta-n-leka=e nir-ked-a ***daɽa-duɽu raʈa-paʈa***
ja juɳu ja bir ja diri-ko-te-ge ***daɽa-duɽu*** *idi-n-ta-n-a-e*

"Then [the demon] left his house, he ran, **daɽa-duɽu** dust flying everywhere, the bushes crackling **raʈa-paʈa**, the forest, stones he is taking with him as he is running **daɽa-duɽu** by himself"

In this utterance, different expressives appear adjacent to each other though they are modifying different actions in the sentence, for instance, *raʈa-paʈa* (modifies *juɳu* 'bush') and *daɽa-duɽu* (modifies *nir* 'to flee'). The depiction is emphasized by the repetition of nucleus vowels a/u throughout the utterance and the 'echo' at the end of the clause. This serialization of expressives adds to the depictive quality by foregrounding the sensual features of the utterance. Another interpretation of this usage would follow Diffloth's (1979) proposal of separate prosaic and expressive modes of language. It has been noted that expressives can exist as their own utterance, not linked morpho-syntactically to other adjacent elements, and not requiring a subject or predicate. In this sense, they could be considered 'sentences' in and of themselves. The expressives in this example make reference to the bush and the running, but evoke the imagery expressively, and without a direct link to the prosaic utterance. Concatenation of two expressives used in this somewhat deviant syntactic context delivers a more nuanced and poetic image.

4. Expressive Elicitation

In addition to storytelling tasks, we also asked Purti to explain the meanings of expressives through conventional elicitation in order to discover specific contrasts (Badenoch 2020). However, because expressives are 'depictions', we asked her to *enact* rather than simply *explain* the meanings. In doing so, we obtain a fuller range of contrasts, through speakers' communication of specificity, intensity, vividness, and euphony (Watson 2001). This approach was not solely our request, but rather a decision to pick up on her 'natural' way of explaining expressives which, more often than not, included gesturing and play-acting.

The following expressives all depict the act of eating. Each has a different type of motion and may involve sound. There may also be a person-type evoked in the use of the expressive. Purti often provides information on the type of person that would typically perform the type of action under discussion (see Choksi 2020).

4.1 lugum-lugum

This expressive is used to depict the chew motion of eating but is also used for someone that mutters when they speak.

baba lad ***lugum-lugum****-ta-n=e jom-ta-n-a*
"Somebody is eating a rice flour pancake **lugum-lugum**."

This expressive was described as a silent chewing motion often used for people who don't have teeth (old people, babies). The lack of teeth was indexed through sunken cheeks and the chewing motion was circular as food is gummed at some length of time in order to soften it. It was keyed in the video through a visual chewing motion with no sound.

4.2 cakob-cakob

Here, the expressive focuses on a motion and a sound.

cakɣab-ge isin-ta-n aɣaʔ ***cakob-cakob****-ta-n jom-oʔ-a*
"Undercooked greens are eaten **cakob-cakob**."

This expressive was enacted through a chewing motion with mouth open and lips making a smacking sound. The smacking sound enacts the effort needed to chew the hard vegetables and indicates that they are not preferred (Badenoch, 2017).

4.3 sar-sor

Using this expressive, attention is drawn to the movement of not only the mouth but the entire head with a different type of sound

sukuri-ko-leka…lolo chij ***sar-sor****-ta-n jom-oʔ-a*
"Like pigs….hot things are eaten **sar-sor**"

Here the mouth is open wide. There is a snorting or sucking sound, and the head moves back and forth so that air can get in to cool the hot food. Here, however, the emphasis is on the unattractive motion of the head.

japani hoɣo-ko ***sar-sor****-ta-n ramen=ko jom-a*
"Japanese people eat ramen noodles **sar-sor**."

Here, she displays a slurping sound. The sound of chewing is less prominent while there is slight snorting and the head moves side to side. Here, we can recognize the idea of eating hot things, as the noodles are slurped to accommodate the heat at which they are served. This use of *sor-sor* is also a perceptive statement on the eating of ramen, as usually, one's

head raises up together with the slurp, as it is preferred that the noodles are not bitten through but rather sucked up in several slurps.

The elicitation of expressives can provide sense contrasts along multiple axes of perception, such as sound (*sar-sor*, *cakob-cakob*) vs silence (*lugum-lugum*), and sight (*lugum-lugum*, *cakob-cakob* vs. *sar-sor*), as well as different types of movement involving the lips, jaw, and head. This type of elicitation can also offer more sample sentences to provide baseline comparative material for analyzing expressive use in narrative or other interactions. Expressive depictions are still context-dependent; the meaning is linked to the contexts created in the elicitation session. Thus, *sar-sor* has different sense foregrounding depending on the situation, one of which is a comparison to a pig, the other is a comparison of a person eating noodles. These fine nuances can be representative of a moral semantics in which Mundari expressives are used (Badenoch, Choksi and Purti, 2019).

5. Conclusion

Generally speaking, expressives have not received much attention from linguists, although recent publications on expressives in the South Asian linguistic area (Badenoch and Choksi 2020; Williams 2020) have started to raise the profile of this area of study. Typically, there has been an idea that they are not 'real language', shared sometimes by both researchers and speakers. This is partially true because the semantics of expressives are complex, as they deal with multi-sensory depictions that are situated in a speakers' direct experience. In our storytelling data, our consultant used a range of tools, including symmetric echoing, gesturing, phonation, and serialization. The elicitation of expressives is notoriously difficult but using a combination of story-telling and speaker enactments, we obtain a better idea of the semantic domains and pragmatic nuances of expressives. Finally, an enhanced understanding of Munda expressives gives us not only a deeper view on the grammar of the language but opens up new perspectives on how language is embedded in cultural practices and world-views.

References

Badenoch, Nathan. "Sticky Semantics: Expressive Meaning in Mundari." In *Expressives in the South Asian Linguistic Area*, edited by Nathan Badenoch and Nishaant Choksi,17-34. Leiden: Brill,2020.

Badenoch, Nathan, and Nishaant Choksi, eds. *Expressives in the South Asian Linguistic Area*. Leiden: Brill, 2020.

Badenoch, Nathan, and Toshiki Osada. *A Dictionary of Mundari Expressives*. Tokyo: ILCAA, 2019.

Badenoch, Nathan, Nishaant Choksi, and Madhu Purti. "Expressives as Moral Propositions in Mundari." *Indian Linguistics* 80, no. 1-2 (2019):1-17.

Badenoch, Nathan, Nishaant Choksi, and Madhu Purti. "Moral expressives in Mundari." Paper presented at the International Conference of the Linguistic Society of India, Mysore, India, December 5-7, 2018.

Choksi, Nishaant. "Expressives and the multimodal depiction of social types in Mundari." *Language in Society* 49, no. 3 (2020): 379-398.

—. "Sense, person, object: diagrammatic networks and the enactment of expressives in Munda." Paper presented at 116thMeeting of the American Anthropological Association, Washington, D.C, November 30, 2017.

Diffloth, Gerard. "Notes on Expressive Meaning." *Chicago Linguistic Society* 8, no.44 (1972): 440–47.

—. "Expressive Phonology and Prosaic Phonology in Mon-Khmer.", In *Studies in Tai and Mon-Khmer Phonetics and Phonology in honour of Eugénie J.A. Henderson,* edited by T.L. Thongkum et al., 49-59, Chulalongkorn University Press, 1979.

Dingemanse, Mark. "Making new ideophones in Siwu: Creative depiction in conversation." *Pragmatics and Society* 5*,* no.3 (2014): 384-405.

Haiman, John. *Cambodian: Khmer*. Amsterdam: John Benjamins Publishing, 2011.

Lahti, Katherine, Rusty Barrett, and Anthony Webster. "Introduction: Ideophones: Between Grammar and Poetry." *Pragmatics and Society* 5, no.3 (2014): 335–40.

Mohan, Shailendra. "Echo-word formation in Hindi". *Indian Linguistics* 67*,* no.1-4 (2006):119-26.

Nuckols, Janis. *Sounds Like Life: Sound-symbolic Grammar, Performance and Cognition in Pastaza Quechua* (Oxford Studies in Anthropological Linguistics) Oxford: Oxford University Press, 1996.

Osada, Toshiki. "Mundari Expressives." In *Austro-Asiatic Linguistics*, edited by KS Nagaraja, 271–82. Mysore: Central Institute of Indian Languages, 2010.

Osada, Toshiki, Madhu Purti and Nathan Badenoch. "Expanding the Model of Reduplication in Mundari Expressives." In *Expressives in the South Asian Linguistic Area*, edited by Nathan Badenoch and Nishaant Choksi, 78-99. Leiden: Brill, 2020.

Phillips, Jacob B., and K. David Harrison. “Munda Mimetic Reduplication.” *Canadian Journal of Linguistics/Revue Canadienne de Linguistique* 62, no.2 (2017): 221–42.

Watson, Richard L. "A comparison of some Southeast Asian ideophones with some African ideophones." In *Ideophones*, edited by F. K. Erhard Voeltz and Christa Kilian-Hatz, 385-405. Amsterdam: John Benjamins Publishing, 2001.

William, Jeffery P., ed. *Expressive Morphology in the Languages of South Asia*. New York: Routledge, 2020.

CHAPTER FIVE

THE PAST SUFFIXES OF HILL KORWA[1]

MASATO KOBAYASHI

1. Introduction

Hill Korwa is a Kherwarian Munda language spoken in a few villages of northern Chhattisgarh.

In one of the first descriptions of Korwa, Grierson [Konow] (1906: 147–166) mentions "Erṅgā or Singlī" as "a slightly different dialect" of Korwa, spoken in Jashpur district by about 500 people (p.149) out of 20,227 total speakers according to the Census of 1891. Since the Plains Korwas speak local Indo-Aryan languages, and since the Hill Korwas in Surguja district appear to have stopped speaking Korwa decades ago,[2] Ernga (also Erngha) is the only Hill Korwa still in active use as far as we know.[3] In the rest of this paper, we refer to Ernga Hill Korwa as 'Hill Korwa'. We visited Kado Pani, a Hill Korwa village in Jashpur District of Chhattisgarh, in February 2014, March 2015, and March 2016, and this is a report of our ongoing descriptive work on the language.

In his survey, Grierson pointed out several features of Hill Korwa (Grierson 1906: 150-151): Productive use of the infix *-nV* as in *ka-na-lom* from *kalom* 'last year'; the progressive suffix *-ta* instead of *-tan*, which is sometimes used for the past; simple past in *-eḍ*, *-aḍ*, *-en* (*-yan*), and *-an*; perfect in *-ted*, *-teḍ*, *-teṛ*, and *-ter* as in *sab-ted-a* {seize-PRF-FIN} 'have seized', perfect in *-ked* and *-ken* as in *yam-keṛ-a* {get-PRF-FIN} 'got', *goc̓-ken-a* {die-PRF-FIN} 'died'; and the suffixes *-led*, *-len* for the remoter past. Grierson concludes that Hill Korwa is closely related to Asuri, based on features such as distal demonstratives with *m-* like *man* 'that'.

There are only a handful of works on Hill Korwa after Grierson (1906). Barker (1953), as a result of the author's fieldwork in Surguja, identified Korwa phonemes and recorded many inflected verb forms. Bahl (1962) demonstrated that Hill Korwa has solidly Kherwarian basic vocabulary,

such as *ayum* ‘to hear’, *jonoʔ* ‘broom’, and *huḍiŋ* ‘small’. Prasad (1985) and George (2014), Hindi-(English-) Korwa phrase books, also contain verb forms.

In this paper, we will discuss the tense or aspect marking of Korwa verbs based on the forms in the narrative texts we recorded (Kobayashi 2015, 2016, 2017).

2. Verb Morphology of Korwa

A Korwa verb minimally consists of a verb root and normally takes suffixes marking the object, tense, aspect, and modality and a clitic showing agreement with the subject.

A verb root may be extended with a causative prefix *a-* or *ece-*, by reduplication or by the reflexive suffix *-n* or *-oʔ*.[4] A verb root serves as an infinitive, e.g. *wec ̃ awei-ta* {come can-NPST} ‘can come’, and also as a noun, as in *jauʔ* which means ‘food’ as well as ‘to cook’.

A verb base is followed by a past or perfect tense-aspect suffix, or the benefactive suffix *-o*/*w* followed by an indirect object agreement marker.[5] If the verb is transitive, a direct object agreement marker comes after that. At the end of these suffixes, the finiteness marker *-a* optionally occurs. Alternatively, the non-past suffix *-tan*/*-ta* occurs at the end of the verb. The imperative marker *=me* may also be attached to a preceding word just like the subject agreement marker, e.g. *datrom=me agu-aṭu* {sickle=IMP bring-leave} “Bring and leave a sickle!”, and we treat it as a clitic. To sum up, the verb forms we have found so far consist of the following five slots.

$_1${RED/*a*/*ece*-ROOT}+$_2${*-n*/*-oʔ*} + $_3${PST/PRF/BEN}+$_4$OBJ+$_5${*-a*/*-tan*}

A direct object marker follows the verb base or the past or perfect suffix as in

(1) *gog-e-a* {take.away-3SG.OBJ-FIN} ‘takes it away’

(2) *jog-e-a* {chase-3SG.OBJ-FIN} ‘chases it’ vs. *joʔ-a* from /jog-ʔ-a/ {chase-PST-FIN} ‘chased’

(3) *ader-ku=me* {get.in-3PL.OBJ=IMP} ‘get them in!’

(4) *aṭu-ter-i-a* {leave-PRF-3SG.OBJ-FIN}‘left him’

(5) *joʔ-kiɖ-iɲ-a* {chase-PRF-1SG.OBJ-FIN} ‘chased me’

(6) *puʈiɖ-kye-m-ta* {tear-SUBJ-2SG.OBJ-NPST} ‘would tear you apart’

When an indirect object is marked, it occurs with the benefactive suffix *-o*/*-w*[6] as in

(7) *kata-o-e-tan* {speak-BEN-3SG.OBJ-NPST} ‘speaks to him/her/it’

(8) *jauʔ-w-alaŋ-a* {cook-BEN-1DU.OBJ-FIN} ‘cooks for us’ (or *jauʔ-wa-laŋ-a*).

Verb roots are combined as a serial verb to denote two consecutive or concurrent actions, just as Pucilowski (2013:98) describes about Ho.

(9) *joʔ-bolo-ter-i-a* {chase-corner-PRF-3SG.OBJ-FIN} ‘chased and cornered it’;

(10) *ɖega-enecˀ-tan* {jump-play-NPST} ‘jumps and plays’.

A few morphemes combine with another root as a vector verb and mark aspect or voice. For example, *-jom* serves as the middle voice marker, as in

(11) *ɲel-jom=me* {look-MID=IMP} ‘watch for yourself!’

-goɖ (*gwoɖ* in Barker 1953) is a completive marker, as in

(12) *gocˀ-goɖ-kye-m-a* {kill-COMPL-SUBJ-2SG.OBJ-FIN} ‘would kill you’

Another interesting innovation of Korwa verb morphology is that the subject marker often becomes a proclitic and forms a phonological domain with the following verb, as with the following verb in clause-initial position.

(13) *kin=kudae-a* {3DU=return-FIN} ‘they come back’.

3. Tense or Aspect Markers

I will now discuss how the tense and aspect are marked on the verb.

3.1 Non-past *-ta/-tan*

-ta is a suffix of the non-past which covers the present and future tenses. It is often found in contexts referring to the future. It occurs with both transitive and intransitive verbs, e.g.

(14) *wec̀-ta* {come-NPST} 'come'
(15) *sen-ta* {go-NPST} 'go'
(16) *raʔ-ko-ta* {call-3PL.OBJ-NPST} 'call them'
(17) *jalm-eɲ-ta* {lick-1SG.OBJ-NPST} 'lick me'

Since *-ta* is often pronounced *-tan*, and since a final /n/ often drops as in *idan~ida* 'be', *-ta* is considered to be cognate with the Kherwarian imperfective *-tan* as in Mundari or Ho. Note that Barker (1953:19) gives *rag-e-tan-a* {call-3SG.OBJ-PROG-FIN} 'I am calling him', a form with *-tan* followed by *-a*.

3.2 Finite *-a*

Korwa has the suffix *-a*, which is also found in Mundari, Ho, and Santali as a finiteness marker. Many verbs are not marked for the tense at all and only have this suffix. Some of them occur with words like *rōj* 'everyday' and imply habitualness, as in

(18) *rōj=e sēn-a* {daily=3SG go-FIN} 'He goes everyday'.

But forms that end in *-a* and contain no tense-aspect suffix such as the following, also occur in past contexts. As these forms occur in narratives, we treat them as cases of the historical present.

(19) *sodor-a* {arrive-FIN} 'arrived' (Kobayashi 2016, 0'35, 3'53)

(20) *iye jaug-e-a* {who cook-3SG.OBJ-FIN} 'Who cooked it?' (Kobayashi 2016, 216)

Some verbs are marked with the completive aspect marker *-goḍ* and *-a*, with no other tense-aspect marker, e.g.

(21) *kata-goḍ-a* {speak-COMPL-FIN} 'spoke'.

Based on these examples, we conclude that *-a* is a tense-neutral finiteness marker occurring in non-future contexts. The suffix *-a* is indistinct after verb roots ending in *a*, such as *tora* 'go', *ida* (*idan*) 'is'.

3.3 Past *-ḍ*

-ḍ or *-eḍ* occurs in past and often anterior contexts, e.g.

(22) *jauʔ-aṭu-ḍ-a* {cook-leave-PST-FIN} 'cooked and left'

(23) *sen-eḍ-a* {go-PST-FIN} 'went'.

Anderson (2007:123) and Anderson et al. (2008:223) gloss *-e* of *-eḍ* as an aspect marker. If it is a separate morpheme, it might be cognate with Mundari 'cislocative' aspect marker *-a* as in *itu-ad-ko-a* {teach-*ad*-3PL.OBJ-FIN} (Osada, 1992:95). Since an independent use of *-e* is not observed elsewhere in Korwa, we analyze *-eḍ* as an allomorph of *-ḍ*. As Anderson comments, it is strange that *-ḍ* occurs in an intransitive verb, for it is a transitive suffix in Ho (Pucilowski 2013:118,158). Pucilowski (2013:122) explains Ho *-eḍ* as an experiential marker in conversation, but Korwa *-ḍ* is probably of a different origin. Korwa *-ḍ*/*-eḍ* often occurs when followed by another verb denoting successive action, and its primary function seems to be to mark anterior action. Pucilowski (2013:161) argued that Ho *-ḍ* and *-n* were reanalyzed as past tense suffixes. The same seems to have happened to Korwa *-ḍ*. The connection of Korwa *-ḍ* and Santhali present tense marker *-et* (Anderson 2007:119) is not clear.

3.4 Perfect *-k*

The suffix *-keḍ* and *-ker* occur in transitive verbs in perfective contexts, as in

(24) *joʔ-kiḍ-iɲ-a* {chase-PRF-1SG.OBJ-FIN} 'chased me'
(25) *sab-keḍ-me-a* {catch-PRF-2SG.OBJ-FIN} 'caught you'
(26) *sab-ker-a* {catch-PRF-FIN} 'caught'
(27) *haṭar-ker-a* {stall-PRF-FIN} 'stalled'.[7]

Since intransitive forms such as

(28) *cʰayman-ki-a* {have. a hunch-PRF-FIN} 'have a hunch'

have no *ḍ* or *r*, *ḍ*/*r* might be analyzable as a transitive marker as in Ho. Note that Korwa also has *-ḍ* (*-eḍ* after a consonant), which, unlike *-ḍ* in Ho, functions as a past tense marker (see 3.3).

3.5 Past *-ter/ -tar*

Like *-keḍ* or *-ker*, the suffix *-ter/-tar* occurs with transitive verbs, such as

(29) *asul-ter-a* {nurture-PRF-FIN} 'was keeping'
(30) *aṭu-ter-i-a* {leave-PRF-3SG.OBJ-FIN} 'left him'
(31) *ṭʰurao-tar-a* {collect-PRF-FIN} 'collected'

Pucilowski (2013:124) analyzes Ho *-taḍ* as a punctual marker. The use of *-ter/-tar* as a past suffix is also found in Kera' Mundari (Kobayashi and Murmu 2008:179). Korwa *-ter/-tar* is always used in past contexts but the difference between *-ter/-tar* and *-keḍ is* difficult to identify in narratives. According to the consultants, *-ter/-tar* and *-keḍ* mark the past and the perfect respectively.

3.6 Past Intransitive *-en/-yan/ -ne/ -na*

The past suffix *-en*/*-yan* and *-ne*/*-na* occur only with intransitive verbs, such as:

(32) *sodor-ne* {arrive-PST} 'arrived'
(33) *durub-ne* {sit.down-PST} 'sat'
(34) *gʰaṭao-en-a* {decrease-PST-FIN} 'decreased'
(35) *wēc-na* {come-PST} 'came'
(36) *nīr-yan* {run-PST} 'ran'

-en occurs as a past marker in Naguri and Kera' Mundari too, such as Kera' *hej̓-en-a* {come-PST-FIN} 'came'. Barker (1953) gives *-yen*, as in *sodor-yen-a* {arrive-PST-FIN} 'he has arrived' (Barker 1953:41). It is probably cognate with Ho *-ya-n*, a past intransitive marker as in *hoba-ya-n-a* 'became' (Pucilowski 2013:116).

Since there appears to be little functional difference, *-ne* is considered an allomorph of *-en-a*. On the other hand, Barker (1953) gives *-nena*, where *-nen* is analyzed as a past progressive suffix, as in *ged-nen-a* 'he

was cutting' (Barker 1953:46), or *dohoy-nen-a* 'he was remaining' (p.47), and *-en* (*-yen*) and *-ne* (*-nen*) might be different morphemes.[8]

3.7 Past *-ʔ*

There is yet another morpheme occurring in past contexts, *-aʔ* or *-ʔ*.

(37) *luɻuŋe-aʔ-a* {winnow-3SG.OBJ-PST-FIN} 'winnowed (rice)'[9]
(38) *joʔ-a /jog-ʔ-a/* {run-PST-FIN} 'ran'
(39) *tora-ʔ=kin* {go-PST=3DU} 'they two went'
(40) *jiya-ʔ-a* {live-PST-FIN} 'lived'
(41) *totomb-aʔ-a* {peck-PST-FIN} 'pecked'.

Since the glottal stop is considered an allophone of /g/ before /ʔ/ as is found in the alternation *jog-e-a* 'chases it' vs *joʔ-a /jog-ʔ-a/* 'ran', the origin of this morpheme should probably be sought in a velar stop. Although we cannot find a cognate of '*g'* as a past suffix in other Kherwarian languages, it might be possible to connect '*ʔ'* with '*k'*, which occurs in perfective morphemes such as Santali, Mundari, and Ho *-aka-* (Anderson 2007:119ff., Anderson, p.c.).

4. Discussion

At first sight, the verb morphology of Korwa looks unique due to the sound changes and grammaticalization it seems to have undergone. A closer comparison with other Kherwarian verbs reveals that Korwa has a set of past suffixes similar to Ho, even though their distribution is more limited. As shown in Table 5-1, Korwa and Ho share *-t* and *-en*/ *-yan* as preterite markers, while they do not seem to have equivalents of the ingressive morpheme *-ja* that Mundari has.[10]

	-ɖ/-d	*-k*	*-t*	*-en/-ne*	*-aka*	*-le*	*-ja*
Korwa	PST	PRF	PST	PST	PRF *-ʔ* ?	—	—
Ho	TR	PRF	PNCT	PST *-ya*	PRF	ANT	—
Mundari	TR	COMPL	PROG	Kera' *-en*	CONT	ANT	INGR
Santali	TR *-t'*	COMPL	—	COMPL	RES	ANT	—

Table 5-1: Tense-Aspect markers of Kherwarian Languages

In addition to the relatively smaller number of slots in verb suffixation, Korwa verb morphology is characterized by the loss of the transitivity contrast by-*ɖ*/-*n*. Instead, -*ɖ* and -*n* are incorporated in tense and aspect suffixes in Korwa. The transitive-intransitive contrast is reduced, and only the originally transitive marker -*ɖ*/-*r* occurs in limited contexts. As far as narratives are concerned, marking of tense and aspect does not appear to be made consistently. We need more research to find out the distinction among the tense-aspect suffixes.

Abbreviations

1: first person; 2: second person; 3: third person; ANT: anterior; BEN: benefactive; COMPL: completive; CONT: continuative; FIN: finiteness; IMP: imperative; INGR: ingressive; MID: middle; NPST: non-past; OBJ: object; PL: plural; PNCT: punctual; PRF: perfect; PROG: progressive; PST: past; Q: question; REFL: reflexive; RES: resultative; SG: singular; SUBJ: subjunctive TR: transitive

Notes

[1] I thank the late Bablu Tirkey and Tetru Oraon for arranging the fieldwork, and the Korwa consultants Pandri Bai and Phaguwa Ram for their kind help. The abbreviations in the gloss are taken from Leipzig Glossing Rules unless otherwise mentioned.

[2] When we visited a Korwa village in Surguja District in 2016, only elderly villagers still remembered a few Korwa phrases, and one of them could tell a short story in Korwa.

[3] It is possible that Korwa is still spoken more widely than we assume. Cf. Census of India 2001, "Abstract of speakers' strength of languages and mother tongues", which gives 34,586 as the number of Korwa speakers.

[4] In this paper, we call a morphologically extended verb root a 'verb base'.

[5] The earlier description tells that Korwa marked either direct or indirect object on the verb, e.g. *owa-d-i-a* 'gave to him'(Grierson 1906:151, segmentation mine) vs. *owa-i=me* 'give him!' (p.150), *haʈiŋ-wa-iŋ=me* {divide-BEN-1SG=IMP} 'divide to me'(p.151). In *haʈiŋ-aɖ-i(ɲ)-y-a* 'gave it to me' (p.153), both direct and indirect objects might be marked, if -*y* is a marker of the animate direct object, here a goat. In present-day Korwa, only the direct object, or the subject in verbs such as *idan* 'be', is marked on the verb, and the indirect object is usually expressed by postpositional phrases with *lagin* 'for', e.g. *am-aʔ lagin* 'for you'.

[6] The benefactive suffix -*o/w* is written -*wa(-d)* in earlier description (cf. footnote 5). It might come from the verb *owa(-d)* 'to give' by grammaticalization and phonological reduction. Alternatively, -*o/w* might be a verbalizing suffix like Kera' Mundari -*o/u*.

[7] -*ka* as in *ece-sab-ka* {CAUS-catch-PRF} 'caught' might belong to this suffix.-*ka* is also found in Kera' Mundari (Kobayashi and Murmu 2008:180).

[8] *-ne* is not always past: *daʔ-kun asejom ero bãɽao-ne* {water-and.so.on beg not be. all. right-PST} 'It will become inappropriate [for the elder brother] to beg [the younger brother's wife] water and so on.' (Kobayashi 2016, 2'47)
[9] If *e* in *luɽuŋ-e-aʔ-a* is an object marker, *-aʔ* shows an exceptional morpheme order.
[10] Barker (1953:44) gives *mer=e cawɽ-ja* 'he didn't step' as a contingentive form, and *-ja* might still exist in Korwa. In our corpus, there are forms with *-ja* which seems to signal question or the hortative mode, such as *do-bu homr-e-ja* {let's.go-1PL ask-3SG.OBJ-Q} 'Let's go! Let us ask him (and get a drum like his)!'.

References

Anderson, Gregory D.S., Toshiki Osada, and K. David Harrison. "Ho and the other Kherwarian Languages." In Munda *Languages,* edited by Gregory D.S. Anderson, 195-255, London & New York: Routledge, 2008.

Anderson, Gregory D.S. *The Munda Verb: Typological Perspectives*. Berlin: Mouton de Gruyter, 2007

Bahl, Kali Charan. *Korwa Lexicon*. Chicago: University of Chicago [mimeograph], 1962

Barker, Phillip. R. *The Phonemes of Korowa*. Seattle: University of Washington [mimeograph], 1953

George, Binzy Joseph. *KorvāVārtālāp: Korwa Language Learner's Phrase Book.*Thiruvalla: Wycliffe India, 2014.

Grierson, G. A. *Linguistic Survey of India*, *Vol. IV. Munda and Dravidian*, Calcutta, 1906

Kobayashi, Masato and Ganesh Murmu. "Kera' Mundari." In *Munda Languages,* edited by Gregory D.S. Anderson, 165-194, London & New York: Routledge, 2014.

Kobayashi, Masato. "Erunga Koruwa go minwa'Jakkaru to rôjin.'"*Tokyo University Linguistic Papers* 36 (2015): e1-e8.

Kobayashi, Masato. "Erunga Koruwa go minwa'Tora no hanayome.'" *Tokyo University Linguistic Papers* 37 (2016): e41-e53.

Kobayashi, Masato. "ErungaKoruwa go minwa'Hotaru no muko.'" *Tokyo University Linguistic Papers* 38 (2017): e75-e83.

Osada, Toshiki. *A Reference Grammar of Mundari.* Tokyo: ILCAA, 1992.

Prasād, Svarṇalatā. *Korvā Vārtālāp Nirdeśikā*. Rāṁcī: Bihār Janjātīya Kalyāṇ Śodh Saṃsthān, 1985.

Pucilowski, Anna. "*Topics in Ho Morphophonology and Morphosyntax."* Ph.D. dissertation, University of Oregon, 2013.

CHAPTER SIX

TYPOLOGICALLY QUIRKY CHARACTERISTICS OF PAST AND PERFECTIVE FORMS IN KHERWARIAN

BIKRAM JORA
AND GREGORY D. S. ANDERSON

1.1 Introduction

In this brief report, we outline some typologically quirky features seen when comparing past and perfective with non-past and imperfective forms in the Kherwarian Munda languages. Due to a lack of basic materials on many known varieties, as well as the fluid nature of some of the ethnonyms, both exonyms and endonyms included like Bhumij, Munda, Kol, etc., it is not clear how many languages and dialects there are in the complex continuum we call Kherwarian. The largest and best known KherwarianMunda language is the now Scheduled (official) language of India, Santali, with over seven million speakers primarily in West Bengal, Jharkhand, and Odisha, but also found in Bangladesh and Nepal as well as widely scattered in India due to the first Santali diaspora of the mid-to-late-19th century due to the unsuccessful Santal Rebellion. Kherwarian includes two other major languages, Mundari and Ho, which are closely related to each other but distinct enough lexically, phonologically, and morpho-syntactically to be considered separate languages; socio-linguistically the two groups are very distinct as well. The number of speakers must exceed 1.5 million for Mundari and 1 million for Ho again mainly in Jharkhand and northern Odisha. Other known Kherwarian varieties but most poorly documented include Korwa, Korowa, Koraku, Asuri, Turi, Mahali, Karmali, Koḍa, Birhor, Bhumij, etc.; see Figure 1.

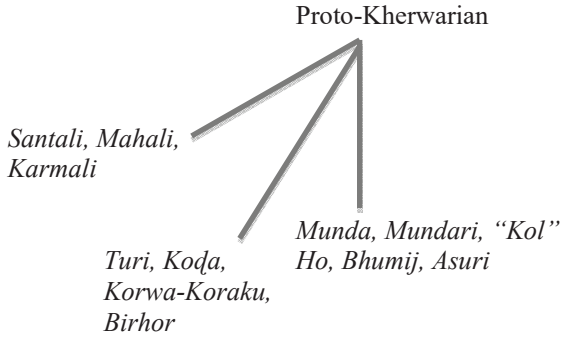

Figure 6-1: One Possible Classification of Kherwarian Lects
(Anderson, this volume)

In section 2 we introduce some characteristically Kherwarian formations that show asymmetries in coding iso-functional and iso-referential participants in copula formations between the present and past forms, where the same participant is encoded either as an object (present) or a subject (past). Section 3 introduces a quirky feature of the Kherwarian language Birhor where past negative forms may be marked by just the negative scope operator that hosts the subject, with all TAM, voice/valence, object agreement, and mood markers suppressed. Section 4 discusses a feature of the perfective series of inflections across Kherwarian where the default perfective/past marker for transitive forms is different than it is for intransitive forms, a feature lacking in the imperfective series of inflections and which results in asymmetries within the perfective series as well such that the more marked members of the perfective series typically occur with both the intransitive/middle marker *–n-* and the transitive/active one in *–d-* in paradigmatic sets (the latter may be suppressed as well in some instances when object marking is present typically. i.e., having another index of transitivity/object referentiality/definiteness, etc.), while the default perfective/past markers for two-place predicates and the one for one-place predicates are specialized in distribution often. Section 5 presents similar asymmetries between the default TAM marker(s) seen in the positive perfective conjugations with those of negative ones in various Kherwarian languages, such that semantic nuances typically associated with those TAM markers in positive forms may be lost when under a negative scope.

1.2 Overview of Inflection in Kherwarian Languages

The Kherwarian verb is characterized by a fair degree of complexity. Since we have data for only a percentage of the Kherwarian languages to date, any reconstruction is necessarily premature and tentative, and thus to be approached with some caution, but that said, one can infer from the data we do have that there were likely two series of inflections in Proto-Kherwarian as there remain today in most, if not all, languages (Anderson and Jora 2018). This contrasted two series of macro-templates of related sets of conjugations, one a perfective series (1)-(2) and one an imperfective series (3)-(4). Within these, one can distinguish positive and negative templates as well as monovalent and polyvalent classes.

(1) PROTO-KHERWARIAN MAXIMAL VERB TEMPLATE [PERFECTIVE SERIES, POSITIVE]

Verb.Stem-APPL-TAM-VOICE/VALENCE-OBJ-IND=SUBJ

or

X=SUBJ Verb.Stem-APPL-TAM-VOICE/VALENCE-OBJ-IND

Where X is any word preceding the verb including the negative scope operator (2)

(2) PROTO-KHERWARIAN MAXIMAL VERB TEMPLATE [PERFECTIVE SERIES, NEGATIVE]

NEG=SUBJ Verb.Stem-APPL-TAM-VOICE/VALENCE-OBJ=IND

Note the order of elements -TAM-VOICE/VALENCE-OBJ with three templatic slots in that order. The imperfective series, on the other hand, showed a different order (3)-(4).

(3) PROTO -KHERWARIAN MAXIMAL VERB TEMPLATE [IMPERFECTIVE SERIES, POSITIVE]

Verb.Stem-OBJ//VOICE-TAM-IND=SUBJ

or

X=SUBJ Verb.Stem-OBJ//VOICE-TAM-IND

(4) PROTO-KHERWARIAN MAXIMAL VERB TEMPLATE [IMPERFECTIVE SERIES, NEGATIVE]

NEG=SUBJ Verb.Stem-OBJ//VOICE-TAM-IND

Subjects are thus encoded by a series of enclitics, typically on the word preceding the verb, while objects are encoded by a series of suffixes which occur directly following the stem in the imperfective series but following the TAM and voice/valence markers in the perfective series.

(5) Ho

aiŋ	*ame=**ŋ***	*nel-**me**-tan-a*
1SG	2SG=**1SG.SUBJ**	see-**2SG.OBJ**-IPFV-IND

'I am looking at you'

(6) Birhor

iŋ	*am=ke*	*ka=**iŋ***	*lel-**me**-kan-a*
1.SG	2.SG=OBJ	NEG=**1SG.SUBJ**	see-**2SG.OBJ**-IPFV-IND

'I am not looking at you'

(7) Santali

am	*iɲ=**em***	*ɖaɽ-oʈo-ki-d-**iɲ**-a*	
2SG	1SG=**2SG.SUBJ**	run-CAUS-TR.PFV-TR/ACT-**1SG.OBJ**-IND	

'You made me run'

(8) Keraʔ Mundari

am	*aiŋ-ke*	*ka=**m***	*kudaɔ-t-**ĩ**-a*
2SG	1SG-OBJ	NEG=**2SG.SUBJ**	run:CAUS-TAM-**1SG.OBJ**-IND

'You didn't make me run'

Note that it is not obligatory for a subject marker to occur dislocated from the verb in Kherwarian, and they may appear at the end of the verb form as well. Nor must the prohibitive negative scope operator always takes the subject clitics either. But these trends for the word immediately preceding the verb to host the subject clitics are strong across the languages of the Kherwarian group.

(9) Birhor

*lel-**me**-kan-a=**iŋ***
see-2SG.OBJ-IPFV-IND=1SG
'I see you'/'I'm looking at you'

(10) Kera? Mundari

aiŋ-ke alɔ *kaj-**iŋ=me***
1SG-OBJ PHB tell-**1SG.OBJ=2SG.SUBJ**
'Don't tell me!'

2.0 Subject and Object Indexing in Kherwarian Copula Forms

We have seen what is 'normal' behavior in the Kherwarian verb, with a subject and object co-referencing in two-place predicates along a nominative-accusative alignment, with some languages treating recipients or addressees as direct or primary objects rather than indirect or oblique objects requiring the applicative (*-a*) following the verb stem, see (11) and (12).

(11) Ho

***a**iŋ* *am* *ka=iŋ* *nel-**a**-me-a*
1SG 2SG NEG=1SG.SUBJ see-APPL-2SG.OBJ-IND
'I won't look at you'

(12) Ho

***a**iŋ* *am* *ka=iŋ* *nel-me-a*
1SG 2SG NEG=1SG.SUBJ see-2SG.OBJ-IND
'I will not see you'

We can now turn to a presentation of some data that are less typical of languages outside of Kherwarian, including other Munda languages, but are characteristic of Kherwarian. This includes the curious shift between marking participants in copular formations as morphological objects in the present but as morphological subjects in the past. So for example, while there is a special form of the negative existential/locational copula in the present for inanimate referents, there is no overt marker for agreement as

there is with animate plurals used in the same constructions, and this takes the object agreement series appearing to the left of the indicative marker, not to its right as with subject markers, compare (13) and (14).

(13) Ho (Deeney 1975: 57)

owaʔ-re	*baba*	*banoʔ-waʔ*
house-LOC	rice	NEG.COP.EXIST-IND(:3SG)

'There is no paddy in the house'

(14) Ho (Deeney 1975: 57)

owaʔ-re	*baŋ-ko-wa*
house-LOC	NEG.COP.EXIST-3PL.OBJ-IND

'They are not at home'

The full set of animate and inanimate forms of the (present) negative existential and locational copula forms in Ho are listed in (15).

(15) Ho (Deeney 1975: 57)

1SG	*baŋ-iɲ-a*		
2SG	*baŋ-me-ya*		
3SG.ANIM[1]	*baŋ-ga-y-a(-)ʔ*	or	*baŋ-ga-ya(-)ʔ*
1DL.INCL	*baŋ-laŋ-a*		
1DL.EXCL	*baŋ-liɲ-a*		
2DL	*baŋ-ben-a*		
3DL	*baŋ-kiɲ-a*		
1PL.INCL	*baŋ-bu-wa*		
1PL.EXCL	*baŋ-le-ya*		
2PL	*baŋ-pe-ya*		
3PL	*baŋ-ko-wa*		
INAN	*banoʔ-wa(-)ʔ*		

Corresponding to the present tense, positive copula forms also take the object series, while past forms of both using the past copula *ʈaiken-a* (negative *ka*=SUBJ *ʈaiken-a*). The present forms take object markers but iso-functional and iso-referential participants in the corresponding past forms are encoded as subjects, and as such are usually indexed on the word immediately preceding the verb, whether it is the negative scope operator or not, compare (16) with (17) and (18) with (19).

(16) Ho (Deeney 1975: 59)
owaʔ-re *menaʔ-**ko**-wa*
house-LOC COP.EXIST-**3PL.OBJ**-IND
'They were at home'

(17) Ho (Deeney 1975: 59)
*owaʔ-re=**ko*** *ʈaiken-a*
house-LOC=**3PL.SUBJ** PST.COP- IND
'They are at home'

(18) Ho (Deeney 1975: 59)
owaʔ-re *ka=**ko*** *ʈaiken-a*
house-LOC NEG=**3PL.SUBJ**PST.COP-IND
'They were not at home'

(19) Ho (Deeney 1975: 57)
owaʔ-re *baŋ-**ko**-wa*
house-LOC NEG.COP.EXIST- **3PL.OBJ**-IND
'They are not at home'

Hasadaʔ Mundari shows a similar pattern (20)-(23).

(20) Hasadaʔ Mundari (Osada 1992: 119)

Soma *oɽaʔ-re* *menaʔ-i-a*
Soma house-LOC COP.EXIST-3SG.OBJ-IND
'Soma is in the house'

(21) Hasadaʔ Mundari (Osada 1992: 54)

parkom *latar-re* *tiʤu-ko* *menaʔ-ko-a*
bed under-LOC insect-PL COP.EXIST-3PL.OBJ-IND
'There are insects under the bed'

(22) Hasadaʔ Mundari (Osada 1992: 120)

hon-ko *oɽaʔ-re* *baŋ-ko-a*
child-PL house-LOC NEG.COP-3PL.OBJ-IND
'Children are not in the house'

(23) Hasadaʔ Mundari (Osada 1992: 60)

du ganʈa sida-re rãʧi-re=ko taiken-a
two hours before-LOC Ranchi-LOC=3PL.SUBJ PST.COP-IND
'They were in Ranchi two hours ago'

Curiously, copula formations used in possessive functions show a similar pattern: animate possessa in present tense forms are treated like objects (24)-(30).

(24) Ho

*aiɲa(ʔ) bəria kuːihon-kin mena(ʔ)-**kin**-a*
1SG:GEN two.ANIM girl.child-DL COP-**3DL.OBJ**-IND
'I have two daughters'

(25) Ho

*aiɲa(ʔ) bəria kuːihon-kin baŋ-**kin**-a*
I:GEN two. ANIM girl.child-DL NEG.COP-**3DL.OBJ**-IND
'I don't have two daughters'

(26) Bhumij

*iɲa(ʔ) bəria kuɽihon-kin mena(ʔ)-**kin**-a*
1SG:GEN two.ANIM daughter-DL COP-**3DL.OBJ**-IND
'I have two daughters'

(27) Bhumij

*iɲa(ʔ) bəria kuɽihon-kin baŋ-**kin**-a*
1SG:GEN two.ANIM daughter-DL NEG.COP-**3DL.OBJ**-IND
'I don't have two daughters'

(28) Keraʔ Mundari

*aiɲa(ʔ) du ʈʰɔ kuɽihɔn hen-**kin**-a*
1SG:GEN two CLSSFR daughter COP-**3DL.OBJ**-IND
'I have two daughters'

(29) Kera? Mundari

*aiɲa(ʔ) du ʈʰɔ kuɽihɔn ka li-**kin**-a*
1SG:GEN two CLSSFR daughter NEG NEG.COP-**3DL.OBJ**-IND
'I don't have two daughters'

(30) Tamaɽia Mundari

*aĩja(ʔ) barija honkuɽi-kin baŋ-**kin**-a*
1SG:GEN two.ANIM daughter-DL NEG-**3DL.OBJ**-IND
'I do not have two daughters'

But these same referents are morphologically encoded like subjects in the past forms, and thus often on the word immediately preceding the verb, including the negative scope element (31)-(35).

(31) Ho (Deeney 1975: 63)

*aliɲ=do hon-ko ka=**ko** ʈaiken-a*
1DL.EXCL=TOP child-PL NEG=**3PL.SUBJ** PST.COP-IND
'We two had no children'

(32) Ho

*aiɲa(ʔ) bəria ku:ihon-kin ka=**kin** taiken-a*
1SG:GEN two.ANIM daughter-DL NEG=**3.DL.SUBJ** PST.COP-IND
'I did not have two daughters'

(33) Bhumij

*iɲa(ʔ) bəria kuɽihon ka=**kin** taiken-a*
1SG:GEN two.ANIM daughter NEG=**3DL.SUBJ** PST.COP-IND
'I did not have two daughters'

(34) Tamaɽia Mundari

*aĩja(ʔ) barija honkuɽi-kin ka=**kin** taiken-a*
1SG:GEN two.ANIM daughter-DL NEG=**3DL.SUBJ** PST.COP-IND
'I did not have two daughters'

(35) Kera? Mundari

aiŋa(ʔ)	*du*	*ʈʰɔ*	*kuʈihɔn*	*ka=**kin***	*dɔhɔnken-a*
1SG:GEN	two	CLSSFR	daughter	NEG=**3PL.SUBJ**	PST.COP-IND

'I did not have two daughters'

Note that as (31)-(32) show, it does not matter if the possessor is encoded by a genitive marked pronoun (32) or an unmarked one (31) in Ho (of course in this example the unmarked one appears with the topic marker *=do*).

Some Kherwarian languages show individual divergences from this pattern. Thus, other Bhumij varieties than the one we recorded show different past tense forms in such copula constructions. In the speech of the Bhumij whose language was the subject of Ramaswami's (1992) study, the past forms instead of using *taiken-a* are formed by suffixing the perfective intransitive series with inanimate subject forms (36)-(37).

(36) Bhumij (Ramaswami 1992: 146)

aiŋ-a	*ʈʰori*	*bano-ʤa-n-a*
1SG-GEN	knife	NEG.COP.EXIST-PFV.ITR/MDL-ITR/MDL-IND

'There was no knife with me'

(37) Bhumij (Ramaswami 1992: 146)

gaɖa-re	*daʔa*	*bano-ʤa-n-a*
river-LOC	water	NEG.COP.EXIST-PFV.ITR/MDL-ITR/MDL-IND

'There was no water in the river'

Curiously, for some Birhor speakers, there is an as yet unexplained split between singular and non-singular negative present copula forms. Non-singular forms (38)-(45) behave like the other Kherwarian languages in these copula forms and encode the participant with the object series.

(38) Birhor (Kameshwar Birhor's speech)

alaŋ	*oɽaʔ-re*	*banuʔ-laŋ-a[ʔ]*
1DL.INCL	house-LOC	NEG.COP.EXIST-1DL.INCL.OBJ-IND

'You and I are not in the house'

(39) Birhor

aliŋ *oɽaʔ-re* *banuʔ-liŋ-a*
1DL.EXCL house-LOC NEG.COP.EXIST-1DL.EXCL.OBJ-IND
'He and I are not in the house'

(40) Birhor

aben *oɽaʔ-re* *banuʔ-ben-a[ʔ]*
2DL house-LOC NEG.COP.EXIST-2DL.OBJ-IND
'You two are not in the house'

(41) Birhor

hinkin *oɽaʔ-re* *banuʔ-kin-a*
3DL house-LOC NEG.COP.EXIST-3DL.OBJ-IND
'They two are not in the house'

(42) Birhor

abu *oɽaʔ-re* *banuʔ-bu-a*
1PL.INCL house-LOC NEG.COP.EXIST-1PL.INCL.OBJ-IND
'We (INCL) are not in the house'

(43) Birhor

ale *oɽaʔ-re* *banuʔ-le-ja*
1PL.EXCL house-LOC NEG.COP.EXIST-1PL.EXCL.OBJ-IND
'We (EXCL) are not in the house'

(44) Birhor

ape *oɽaʔ-re* *banuʔ-pe-ja*
2PL house-LOC NEG.COP.EXIST-2PL.OBJ-IND
'You are not in house'

(45) Birhor

hinku *oɽaʔ-re* *banuʔ-ku-a*
3PL house-LOC NEG.COP.EXIST-3PL.OBJ-IND
'They are not in the house'

Singular forms on the other hand (46)-(48), add an (animate?) element *-iʄ-* following the negative copula form to which then attaches the indicative suffix *–a* and then the subject clitics for 1SG, 2SG, and 3SG.

(46) Birhor (Kameshwar Birhor's speech)

iŋ	*oɽaʔ-re*	*banuʔ-iʄ-a=ŋ*
1SG	house-LOC	NEG.COP.EXIST-ANIM-IND=1SG.SUBJ

'I am not in the house'

(47) Birhor

am	*oɽaʔ-re*	*banuʔ-iʄ-a=m*
2SG	house-LOC	NEG.COP.EXIST-ANIM-IND=2SG.SUBJ

'You are not in the house'

(48) Birhor

hini	*oɽaʔ-re*	*banuʔ-iʄ-a=e*
3SG	house-LOC	NEG.COP.EXIST-ANIM-IND=3SG.SUBJ

'He is not in the house'

How these quirky systems arose in Proto-Kherwarian or Proto-North Munda all remains a mystery, much less the divergent development that Birhor shows, but we hope that future research will shed some light on this. Albeit odd, nevertheless, the original copula agreement patterns are quite robust and found across the Kherwarian languages.

3.0 TAM-less Forms in the Birhor Past

Typically the verb in the Kherwarian languages has many morphemes encoding the range of elements seen in the templates and exemplified in section 1 above, and the Santali form in (49). However, unlike its sister languages, Birhor stands apart in its predilection to the suppression of TAM marking in past forms, along with other grammatical markers, in negative past forms. The negative scope operator usually appears with the subject clitic, but the verb appears in an otherwise unmarked form (50)-(52).

(49) Santali

am	*iŋ*	*ba=m*	*ɖaɽ-oʧo-li-d-iŋ-a*
you	I	NEG=2SUBJ	run-CAUS-TR.ANT-TR/ACT-1OBJ-FIN

'You didn't make me run'

(50) Birhor

sukri=ʈa	*ale=ke*	*ka=i*	*lel*
pig=DEF	1PL.EXCL=OBJ	NEG=3SG.SUBJ	see

'The pig did not see us'

(51) Birhor

am	*hɔɽ=ke*	*kə=m*	*gɔʧ*
2SG	man=OBJ	NEG=2SG.SUBJ	kill

'You did not kill the man'

(52) Birhor

iŋ	*am=ke*	*ka=iŋ*	*lel*
1.SG	2.SG=OBJ	NEG=1SG.SUBJ	see

'I didn't see you'

4.0 Valence Asymmetries in TAM Marking

Another quirky feature of Kherwarian verbal systems seen in the perfective series is the unexpected specialization of some TAM markers to occurring only with stems that are either monovalent or polyvalent. Across the Kherwarian languages, there is an observable tendency to see the TAM marker *–ke-~-ki-~-k-* to be way more common with transitive predicates than with intransitive ones (53)-(55).

(53) Tamaŗia Mundari

kula	*sukri=ke*	*ka=i*	*goiˀ-k-i-a*
Tiger	pig=OBJ	NEG=3SG.ANIM.SUBJ	kill=PFV.TR-3SG.OBJ-IND

'The tiger did not kill the pig.'

(54) Ho

aiŋ	*hoˀ=ke=ŋ*	*goiˀ-k-i-a*
1SG	man=OBJ=1.SUBJ	kill-PFV.TR-3.OBJ-IND

'I killed the man'

(55) Santali

am	*iŋ=em*	*ɖaɽ-oʈo-ki-d-iŋ-a*
2SG	1SG=2SG.SUBJ	run-CAUS-TR.PFV-TR/ACT-1OBJ-IND

'You made me run'

Conversely, there is a massive skewing of the use of the TAM marker *–ja- ~ -ʤa-* (or more specifically *–ja-n- ~ -ʤa-n*) with intransitive and middle predicates (56)-(59).[2]

(56) Ho

koto	*ka*	*rəpuɖ-jə-n-a*
Branch	NEG	break-PFV.ITR/MDL-ITR/MDL-IND

'The branch did not break'

(57) Santali

ɖɛr	*rapud-e-n-a*
branch	break-PFV.ITR/MDL-ITR/MDL-IND

'The branch broke'

(58) Bhumij

koto	*ka*	*rəpud-ʤa-n-a*
branch	NEG	break-PFV.ITR/MDL-ITR/MDL-IND

'The branch did not break'

(59) Birhor

hini	*bir-te*	*kula*	*lel*	*sen-ʤa-n-a=e*
3SG	forest-ALL	tiger	see	go-PFV.ITR/MDL-ITR/MDL-IND=3SG.SUBJ

'He went to the forest to see the tiger'

These two thus form a complex TAM+voice opposition that distinguishes this default past forms from more marked formations like the anterior in *–le-n: -le-d-*(60)-(61) and the perfect in *–aka-n-: aka-d-*(62)-(63) where the same TAM markers stand and used with the separate voice markers *–n-* and *–d-* in opposition to each other,

(60) Ho (Burrows 1915 [1980]: 50)

giti-le-n-a=iŋ	*em-le-n-a=iŋ*
sleep-ANT-ITR/MDL-IND=1SG.SUBJ	give-ANT-ITR/MDL-IND=1SG.SUBJ
'I had slept'	'I had been given'

(61) Ho (Burrows 1915 [1980]: 47)

águ-le-d-á=iŋ	*águ-le-d-iŋ-á=e*
bring-ANT-TR/ACT-IND=1SG.SUBJ	bring-ANT-TR/ACT-1SG.SUBJ-IND=3SG.SUBJ
'I had brought (it)'	'He had brought (it to) me'

(62) Ho (Burrows 1915 [1980]: 71, 49)

em-aka-n-á=iŋ	*giti-aka-n-á=iŋ*
give-PRF-ITR/MDL-IND=1SG.SUBJ	sleep-PRF-ITR/MDL-IND=1SG.SUBJ
'I had been given'	'I have slept'

(63) Ho (Burrows 1915 [1980]: 44)

em-aka-d-iŋ-á=e
give-PRF-TR/ACT-1SG.OBJ-IND=3SG.SUBJ
'He has given (it to) me'

Although sometimes the 'voice/valence' markers express more complex semantic distinctions than just the presence or absence of an object, such as animacy, affectedness, and so on; compare the following Santali forms in this regard:

(64) Santali

iŋ	*iŋaʔ*	*ʤoʤom*	*ti*	*ba=iŋ*	*arub-aka-n-a*
I	I:GEN	~.eat>right	hand	NEG=1	wash-PRF-ITR/MDL-FIN

I didn't wash my right hand.

(65) Santali

iŋ	*gidra*	*ba=iŋ*	*arub-aka-d-ko-a*
I	baby	NEG=1SUBJ	wash-PRF-TR/ACT-3PL.OBJ-IND

I haven't washed the babies.

Here the distinction is one of animacy and affectedness between the use of *–d-* in (61) and the *–n-* in (60), as the verbs the same and thus the same transitivity, so the valence of the predicate or the demotion or suppression of one have nothing to do with the selection of *–n*-vs. *–d-* in such Santali forms.

Similar issues arise when looking at the following examples from Hasadaʔ Mundari as well. The form in (62) is a straightforward causative construction and thus the use of *–d-* makes sense as the form is clearly transitive.

(66) Hasadaʔ Mundari (Osada 1992: 90)

hoɽo-ko =m rasika-rika-aka-d-ko-a
person-PL=2SG.SUBJ rejoice-CAUS-PRF-TR/ACT-3PL.OBJ-IND
'You have made the people rejoice'

In forms lacking the causative, however, both the *–n-* and the *–d-* can be used with basically the same lexical meaning (63)-(64).

(67) Hasadaʔ Mundari (Osada 1992: 90)

hoɽo-ko rasika-aka-d-ko-a
person-PL rejoice-PRF-TR/ACT-3PL.OBJ-IND
'They have rejoiced'

(68) Hasadaʔ Mundari (Osada 1992: 90)

hoɽo-ko rasika-aka-n-a-ko
person-PL rejoice-PRF-ITR/MDL-3PL.OBJ!-IND
'They have rejoiced'

Note that (63) has object markers that encode the logical subject that is typical of experiencer subjects in Kherwarian languages, but the notional subjects are encoded as the subject in (64). The difference between the two is one of affectedness, but this distinction is hard for speakers to translate and textual counts have not been done to determine what factors of orientation or salience trigger the use of subject vs object markers in such experiential constructions.

5.0 Negative-TAM Interdependencies

One last typologically unusual feature found in Munda is that some TAM markers change their meaning under negation. Thus, while as mentioned above, the transitive perfective in most Kherwarian languages (and possibly already so at the Proto-Kherwarian stage) had specialized TAM + voice marker combinations in the unmarked past/perfective reading

(65, repeating 7, 49 above), e.g. *-ki-d-/-ke-d-* in Santali. With negatives, however, as in the Santali example (in 66, repeating 55 above) shows, there is a preference for using the TAM marker *–le-/-li-* instead which functions as either a pluperfect or an anterior marker in positive conjugations but is the default past/perfective TAM marker under negation in our Santali corpus.

(69) Santali

am	*iɲ=em*	*ɖaɽ-otʃo-ki-d-iɲ-a*
you	I=2SUBJ	run-CAUS-TR.PFV-TR/ACT-1OBJ-IND

'You made me run'

(70) Santali

am	*iŋ*	*ba=m*	*ɖaɽ-otʃo-li-d-iɲ-a*
you	I	NEG=2SUBJ	run-CAUS-ANT//NEG-PFV-TR/ACT-1OBJ-IND

'You didn't make me run'

6.0 Summary

Kherwarian Munda languages show a range of quirky features in their past or perfective forms. In past copula formations, in either existential/locational or possessive functions and constructions, the animate participant or referent is encoded as a subject in the past but as an object in the present in negative and positive forms alike. While most Kherwarian languages show elaborate morphology in virtually all verbs, polyvalent negative aorist forms in Birhor show a subtractive pattern where all morphology on the verb itself is suppressed, only subject clitics that appear on the negative scope operator before the bare verb stem is used. Already by the Proto-Kherwarian stage do we see a specialization of past/perfective transitive/active in **-ke-d-* but intransitive/middle in **-ya-n-*. Also, in various Kherwarian languages, an emergent positive vs negative opposition is emerging in the use of TAM markers in positive and negative constructions, such that the now general negative past/perfective marker *–le-* may lose the anterior or pluperfect meaning it typically conveys in positive formations. Some of these features have analogs in other Munda languages like Sora or Korku and thus likely reflect retentions of earlier features. Unraveling the history of such developments is the subject of ongoing research.

Abbreviations

1	1st person	IND	Indicative
2	2nd person	IPFV	Imperfective
3	3rd person	ITR	Intransitive
ACT	Active	LOC	Locative
ANIM	Animate	MDL	Middle
APPL	Applicative	NEG	Negative
AUX	Auxiliary	OBJ	Objective
CAUS	Causative	PFV	Perfective
CLSSFR	Classifier	PHB	Prohibitive
COP	Copula	PL	Plural
DEF	Definite	PRF	Perfect
DL	Dual	PST	Past
EXCL	Exclusive	SG	Singular
EXIST	Existential	SUBJ	Subject
GEN	Genitive	TAM	Tense-Aspect-Mood
INAN	Inanimate	TOP	Topic
INCL	Inclusive	TR	Transitive

Notes

* We would like to thank OpinoGomango for help in the collection of some of this field data. All data used in this presentation, unless otherwise cited, come from the authors' field notes.

[1] Note that this of the pattern for the animate forms ba*ŋ*-OBJ-IND except 3rdsg which is *baŋ-ga-y-aʔ,* where the *–y-* in the 3rdsg could either be an epenthetic glide as it is in other person forms ending in unrounded vowels or the regular object marker of the 3SG. The final *ʔ* here is found only in third singular (inanimate or animate) forms, possibly a reflex of 3SG.SUBJ marker, but these would be unusual in such forms where the referent's existence or presence (or lack thereof) is predicated on is encoded by the object series, as well as the fact that subject marking with inanimate subjects is not typical as a whole in Ho.

[2] So tied together are *ʤa-* and *–n* to being the perfective intransitive/middle form in opposition to the *ke-d/-ki-* form for perfective transitive/active, that this has freed the corresponding transitive/active form to fill a perceived gap in the imperfective series contrasting monovalent and polyvalent inflections with the TAM marker as in the perfective series in Hasadaʔ Mundari (Osada 1992). Thus we find forms that are formally of this perfective series shape with the transitive/active marker and object agreement markers following the TAM marker but with a clearly imperfective type of meaning

(a) Hasadaʔ Mundari (Osada 1992: 100)
seta=ŋ *lel- ʤa-ʔ -i-a*
dog=1SG.SUBJ see-TAM-TR/ACT-3SG.OBJ-IND
'I am looking at the dog'

(b) Hasadaʔ Mundari (Osada 1992: 120)
lel-ʤa-d-me-a=ŋ
see-TAM-TR-2SG.OBJ-IND=1SG.SUBJ
'I am looking at you'

Note that the original form exists too, and like any Kherwarian verb-object agreement can be suppressed as in the following example

(c) Hasadaʔ Mundari (Osada 1992: 99)
seta=ŋ *lel-tan-a*
dog=1SG.SUBJ see-IPFV-IND
'I am looking at the dog'

References

Anderson, G. D. S. this volume. "On the Role of Areal and Genetic Factors in the Development of the Word-Structure and Morphosyntax of the Munda languages".

Jora, Bikram, and Gregory, D.S.Anderson. " Proto- Khewarian Negation, TAM and Person Indexing Interdependencies." In *Austroasiatic Syntax in Areal and Diachronic Perspective*, edited by Mathias Jenny, Paul Sidwell, and Mark Alves,236-257, Leiden: Brill,2020.

Burrows, L. *A Grammar of the Ho language. An Eastern Himalayan Dialect.* New Delhi: Cosmo, 1915 [reprint 1980].

Deeney, J. *Ho Grammar and Vocabulary*. Chaibasa: St. Xavier Publications, 1975.

Osada, T. *A Reference grammar of Mundari.* Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, 1992.

Ramaswami, N. *Bhumij Grammar*. Mysore: Central Institute for Indian Languages, 1992.

CHAPTER SEVEN

PHRASAL AFFIXES AS CLITICS IN THE MUNDA LANGUAGES

ANISH KOSHY

1.0 Introduction

The term 'clitic' (from Greek *kli:no* 'lean') is used for different kinds of formatives in languages, especially when it is difficult to classify such formatives as independent words or as derivational or inflectional affixes. Among the suggested bewildering array of definitions and diagnostics for clitics in the literature, one is likely to find that there is no one definition or a negotiated common understanding of what constitutes or does not constitute a clitic. The extreme theoretical positions on clitics and definitional disagreements are more likely to overwhelm the researcher than give clear yardsticks for the identification of clitics.

According to one of the common definitions of a clitic available with the SIL web portal, "A clitic is a morpheme that has syntactic characteristics of a word, but shows evidence of being phonologically bound to another word." The features of a possible clitic include one or many of the following:

- Phonologically bound but syntactically free
- Functions at the phrase or clause level
- Cannot be integrated into standard discourse without being bound to some other form
- Often has grammatical rather than lexical meaning
- Belongs to closed classes like pronouns, prepositions, auxiliary verbs, and conjunctions
- Usually attaches to the edges of words, outside of derivational and inflectional affixes
- Often attaches to several syntactic categories of words such as head noun, non-head noun, preposition, verb, or adverb
- Phonologically unstressed

2.0 Clitics in Munda languages

Almost all languages have clitics. But they remain mostly unrecognized, at least not to the extent that words and affixes are recognized. This is primarily so, because of their individual characteristics that make them resemble words in some contexts and affixes in some others. Clitics in Munda languages have been recognized for a while now, and it is heartening to note that different scholars use different definitional properties of clitics to identify them.

In sentences like (1) from Mundari, the subject agreement markers in both the clauses are clearly identified as clitics because of their property to phonologically attach to any word or different words within a phrase or a clause. This is the typical behaviour of clitics in many languages.

(1)

ranci-te=ɲ	*sen-ke-n-re*	*sinema=ɲ*	*lel-ke-d=a*
ranchi-OBLQ=1SG	go-AOR-ITR-LOC	cinema=1SG	see-AOR-TR=DEF

'When I went to Ranchi, I saw the cinema'
(Osada 1992: 121)

The kind of clitics that this paper explores is the type of clitics, noted in Peterson (2008) for Kharia. In Kharia, some bound elements are said to "attach to the right-most element of the phrase, regardless of its status, whether lexical or genitive attribute" (441) when their lexical hosts are dropped or not known, as in (2).

(2)

aḍi=yaʔ	*gʰol*	*beŧa=ḍom=ki=te*
ANAPH=GEN	ten	son=3POSS=PL=OBLQ

'His ten sons'
(Peterson 2008: 441)

Examples (1) and (2) present very contrasting approaches to the diagnostics and study of clitics. In cases like (2), the bound elements are actually seen not as attaching to stems but rather as attaching to the entire phrase. This idea of analyzing a set of bound elements as phrasal affixes and not as word-level affixes is very promising with respect to diagnosing clitics in languages. This is in keeping with scholarly traditions that have looked at clitics and affixes at a level deeper than a mere surface level choice of hosts. This methodology analyses clitics more on the basis of the level of attachment than the choice of host for attachment. As phrasal affixes, clitics are those bound elements which are adjoined to syntactic

phrases and are phonological reflexes of morphological rules that do not apply to word stems, while affixes are those which are adjoined to words (Anderson (1992)). In this regard, clitics are not necessarily always bound forms of free words in the language. Phrasal affixes are seen not as inflected lexical items but rather as markers, which are attached to phrases to express inflection. Their positions are determined by rules of phrase-level morphology, qualifying them to be special clitics.

In this paper, I would like to present examples from different Munda languages of attachment of different bound elements that must be treated as phrasal affixation, that is, as clitics. The paper argues that apart from the agreement clitics that choose their hosts indiscriminately (with no regard to their syntactic categories), like in Mundari and Santali, a large number of clitics in the Munda languages are those which are not recognizable by that criterion alone. In many Kherwarian languages, like Mahali, Karmali, Turi, and Bhumij, the subject enclitics may appear only at the end of the verb complex yet are to be analyzed as clitics due to their status as phrasal affixes. The examples (3) and (4) are from Mahali and Karmali, respectively, where the agreement marker is treated as a clitic because of its phrasal attachment.

(3) *em-a-d-iɲ=me*
give-BEN-TR-1=2
'You gave me (it)'
(Grierson 1906: 78)

(4) *meta-ke-t-ku=e*
tell-ASP-TR-PL=3
'He told them'
(Grierson 1906: 73)

3.0 Phrasal Affixes as Clitics

The idea of treating phrasal affixes as clitics, as is argued in this paper, has not always met with universal acceptance. For example, Zwicky (1977, 1987) insists on making a distinction between *clitics* and *phrasal affixes*. This distinction is proposed on the basis of a difference in interaction that is noted to happen between a host and a clitic and between a host and a phrasal affix. According to Zwicky (1977, 1987), a clitic is only prosodically associated with its host whereas a phrasal affix shows morphological interaction with its host. This analysis had even ruled out the English possessive [-s] as a clitic on this count. This position, however,

is not supported by most instances of what are recognized as clitics in the literature. This erstwhile understanding of clitics as only being prosodically associated with their hosts and as being blind to the morphological properties of their hosts stands revised in most analysis of clitics today. Many recent works like Klavans (1985) and Anderson (1992) take the view that clitics are phrasal affixes. Klavans (1983) maintains that for clitics, the actual phonological host can belong to any word class as long as the dominating phrase belongs only to a particular class. Phrasal affixes have also been referred to in the literature as *lexical clitics* (Halpern, 1995), and are characterized as elements, which have "the distribution of a clitic but the morphology and/or phonology of an affix". In these works, clitics are seen as playing grammatical roles and sometimes even having lexical content.

Clitics, especially pronominal and verbal, have been called phrasal affixes owing to the scope properties of these elements. Clitics, as opposed to affixes, have been observed to take a phrase and not a word as its domain of scope. Anderson (1992), for example, notes that clitics and affixes have similar behaviour and the only difference between them is that clitics are phrasal affixes, that is, they are adjoined to syntactic phrases and are phonological reflexes of morphological rules that do not apply to word stems, while affixes are adjoined to words. Clitics are analyzed as bundles of morpho-syntactic features, which are added to the heads. Within the framework of A-morphous morphology, Anderson (1992) introduces the category of DEPENDENT features whose value is assigned to the phrase and transmitted to all its daughters. According to this framework, phrasal affixes are not inflected lexical items but rather agreement markers, which are attached to phrases to express inflection. Their position is determined by rules of phrase-level morphology, which qualifies them to be special clitics in the sense of Zwicky (1977). In this proposal, morpho-syntactic representations have no internal structure at all and morphemes interpret the morpho-syntactic features provided by the syntax, rather than contributing them, thus, positing that morphology converts syntax into phonological strings. This makes inflectional morphology to be the morphology of syntax as it involves grammatical categories that play an important role in the syntax of a language. The terminal nodes of syntactic structures are treated as bundles of morpho-syntactic features that are not phonologically instantiated.

Special clitics are understood to be those whose position within some phrasal unit is determined by principles other than those of the non-clitic syntax (Anderson 1992: 201-202). Thus, special clitics can be seen as "material introduced into Phonological Form by rules of phrasal affixation

entirely parallel to the introduction of affixes within words by Word Formation Rules" (Anderson 2005: 75). This makes special clitics an "overt manifestation of a class of 'Word Formation Rules' that operate on phrases" (81), that is, special clitics constitute the morphology of phrases. Following Anderson (1992, 2005) clitics can be characterized on the basis of three parameters:

- Its SCOPE: the clitic is located in the scope of some syntactic constituents which constitutes its domain
- Its ANCHOR: the clitic is located by reference to the {FIRST vs LAST vs HEAD} element of the constituent in which it appears
- It's ORIENTATION: the clitic {PRECEDES vs FOLLOWS} its anchor.

Since these same parameters can be used to define affixes and clitics, Anderson (2005) proposes clitics and affixes to "belong to a single unified class of rules" (77). This makes the theory of clitics integrally related to issues of phrasal assignment of inflectional properties like tense, case, etc., that is, the application of morphological spell-out rules to a phrase (Anderson 1992, 1993). It is demonstrated that clitic positioning can be described in terms of the interaction of a small set of conflicting alignment constraints, an approach inspired by the tenets of optimality theory. Anderson (2005) contends that the presence of a corresponding free form should not be always taken as being mandatory to declare accent-less forms as being clitics. Considering that in many languages, clitics are often created by the operation of phonological rules on existing forms. The possibility of considering clitics as a primitive like other lexical categories is therefore ruled out. A two-way classification of clitics as phonological clitics and morpho-syntactic clitics is argued for. Anderson's (2005)'s extensive comparison of clitics with affixes demonstrates that clitics are at best described as belonging to morphology, and are neither syntactic objects nor phonological. They constitute specifically the morphology of phrases and hence can be called as "phrasal affixes" (83).

3.1 Clitics in the form of Phrasal Affixes in the Munda Languages

3.1.1 Agreement markers

In the North Munda Kherwarian languages (Mundari, Santali, and Ho), both subject and object are indexed. However, Korku only shows object

agreement. The subject is not necessarily realized by an affix to the verb complex but also by subject markers enclitic to the word immediately preceding the verb. When appearing on the verb complex, it follows the (DE-) FINITIZER suffix becoming enclitic to the entire verbal complex that is, as a phrasal affix.

Anderson and Harrison (2008) talk of a series of subject enclitics (589) used in Remo (as they say, it is in Kharia and Gutob). The subject enclitic (or suffix) forms are as follows:

	SINGULAR	DUAL	PLURAL
FIRST	*-niŋ, -iŋ*	*-naŋ*	*-nay*
SECOND	*-no*	*-pa*	*-pe*
THIRD	*-Φ,-ga*	*-Φ,-ga*	*-Φ,-ga*

Table 7-1 Subject enclitic/suffix forms in Remo

As can be seen, there is no difference between the SINGULAR, DUAL, and PLURAL forms of the 3rd person. Anderson and Harrison (2008)'s motivation for calling them enclitics is not very clear, because they also refer to them as 'suffixes' in brackets. In all the examples, the agreement markers come with the verb as their host. The possible motivation is once again the argument of them being phrasal affixes, in a kind of long-distance hosting, for they can never come immediately after the root and are separated by other inflectional affixes like the TAM categories and hence come at the end. Thus the only fixed position they have is that they are the last elements of a predicate. Some examples of their use are as follows:

(5) *sap-gə-tə-nay*
come-PST.II-N.PST.II-1PL
'We all have come.'
(Fernandez 1968: 26)

(6) *ju-to*
see-N.PST.I
'He/she sees'
(Fernandez 1968: 25)

(7) *bondagada=na remo-le uli sum-to*
Bondagada=GEN person-PL mango eat-N.PST.I
'The people of Bondagada eat mango[es]'
(Anderson and Harrison 2008: 589)

(8) *bəba* *ɖen-t-iŋ*
RDPL:slap PROG-NPST-1
'I am slapping'
(Fernandez 1968: 35, 54)

(9) *a-sum* *ɖen-gi-t-iŋ*
NEG-eat PROG-PST.ITR-NPST-1
'I have not been eating'
(Fernandez 1968: 58)

In Gtaʔ, the subject agreement seems to be prefixal/ proclitic on the predicate. It can come prefixed to the verb stem or be separated from it by the NEG. The NEG itself has to be understood as a phrasal affix (if it is bound) in any construction due to its semantic scope not just over the verb but the entire predicate represented by the verb and its arguments. The following examples demonstrate the same:

(10) *bɽo-k[e]-ne* *dapŋe* *heʔ-baŋ* *nœŋ* *ljo* *ceʔmwa*
fear-T/A-NF immediately today-ABL 1SG field grass
na=big-ɖe *n[a]=a=big* *ɖak-ce* *basoŋ-ke*
2=sow-or 2=NEG=sow QUOT say-T/A
'Fearing he said: From today will you or will you not sow my field?'
(Anderson 2008: 686 [Mahapatra and Zide, no date, I. 26])

3.1.2 Evidence from Incorporation

Incorporation also raises a very pertinent question on the relatedness of elements within a complex word form. That is, if we look at the following examples from Sora, one would notice that with the nominal incorporated, it appears between the verb-head and other verbal inflectional affixes, which would have been attached directly to the verb, had the noun not been incorporated. But with the noun incorporated, these verbal inflectional affixes now appear in close proximity to the incorporated nominal rather than the verb whose inflectional affixes these are. Unless, one posits that these are actually phrasal affixes, and thus clitics, one would be hard-pressed to explain how these verbal affixes still could modify the verb with a nominal in between, with which they have no relationship. If one were to accept this argument, then a host of verbal affixes would have to be treated as phrasal affixes, that is, as clitics, including agreement markers, same subject markers, intransitive markers, the infinitive markers, etc. The only contrary position one could take is

that the incorporated nominal is invisible to the affixes, and therefore the integrity of the word headed by the verb remains unaffected even post incorporation, which is a difficult position to establish. Verbal functional elements, separated from the verb head by an incorporated nominal, are analyzed to be in a long distance hosting relationship with the head. Therefore, they are treated as clitics, which also establishes their presence as bound markers, even in non-incorporated structures, as a kind of loose bonding, leading to a possible clitic analysis for them in all structures.

(11) *le:m-si:=t=am*
bow-hand=NPST=2
'I bow to your hands'
(Ramamurti 1931: 43-44)

(12) *ɲam-ɟo:=le=n*
catch-fish=NF/SS/PST=ITR
'Having caught fish'
(Ramamurti 1931: 142)

(13) *ə-gik-kid=ben*
2PL-see-tiger=INF
'(for you) to see the tiger'
(Ramamurti 1931: 44)

The following sentences from Sora, show not only the incorporation of an argument, object in the first case, and subject in the latter; they also show how inalienable possession is marked with the help of enclitic pronouns, which when the possessor is incorporated appear enclitic to the verbal complex itself. No agreement markers appear in the verbal complex.

(14) *soi-tam=t=am*
burn-mouth=NPST=2
'I will burn your mouth'
(Ramamurti 1931: 142)

(15) *kuŋ-bəb=t=əm*
shave-head=NPST=2
'Your head is shaven'
(Biligiri 1965b: 240)

In verb complexes with incorporated subject and/or object and serialized verbs, the person and number markers always appear in the final position:

(16) *paŋ-ti-dar=iɲ=te:n*
bring-give-cooked.rice=1=3.PST
'He brought and gave me cooked rice'
(Ramamurti 1931: 43)

The flexibility of the PST and NPST markers in this sentence is quite interesting. Do these markers have the flexibility to come after the 1st verb, both verbs, or at the end? Is this flexibility possible for an inflectional affix or only for a phrasal affix? And what does the flexibility or the random appearance and non-appearance of agreement markers tell us about their status as normal or phrasal affixes? When talking of the order of elements, Anderson (2007: 193-94) says that "the order of elements in serialized and incorporated sequences is mostly set in Sora…", but also adds that "… the intransitive marker may appear following either $verb_1$ or $verb_2$." How does one explain such flexibility in the positioning of elements within a word unless one assigns them the categorical status of clitics, which would then explain the freedom of movement/placement that these bound affixes retain/show. The fact that the bound markers representing GNP and TENSE, appear only once but have scope over both the compounded verbs, is taken as very clear evidence of the affixation being phrasal. This argues for treating both the GNP and the TENSE markers in the language as clitics.

(17) *gil=le-ɟir=iɲ*
see=PST-leave=1
'See me before you go'
(Ramamurti 1931: 44)

(18) *ti-ɟum=t=am*
give-eat=NPST=2
'I'll give you to eat'
(Ramamurti 1931: 44)

3.1.3 Evidence from Compound Verb Constructions

In compound verb constructions in **Mundari**, the agreement clitic appears on the preceding element to the compound verb and only once. The definitizer clitic also appears at the end of the compound verb form

and only once. The TAM categories also appear only once, as in the following:

(19) *ne gaɽa poʈpoʈia=te=ko har-parom=ke=d=a*
this river mortorbike=INSTR=3PL drive-cross=COMPL=TR=DEF
'They drove the motorbike and crossed the river'
(Osada 2008: 136)

Unless we are to conceptualize 'driving a motorcycle and crossing a river' as a single event, we have to look at the above structure as involving two different verbs but modified only by a single set of inflectional markers of ASPECT, TRANSITIVITY, and FINITENESS. This is possible, as is discussed in the literature on clitics, only if these bound markers are treated as having scope over both the coordinated elements by being attached phrasally. That is, these bound markers have to be treated as phrasal affixes, that is, as clitics.

In Gutob too, in a compound/complex verbal predicate [For example, *'Beat up and come back'; 'go and come'*, etc.], the pronominal subject agreement appears enclitic and after the TENSE marker either on the first verb or the second verb. This is clearly another instance of an affix being attached phrasally with scope over multiple verbs. Thus, the agreement marker is clearly established as a phrasal affix, that is, as a clitic.

(20) *ɟom=lai niŋ bu-oʔ pi=loŋ=niŋ*
Jom=ACC I beat.up-PST:TR come-FUT:ITR=1
'I will beat up Jom and come back'
(N Zide 1997: 316)

(21) *ɟom=lai bu-oʔ=niŋ pi=loŋ*
Jom=ACC beat.up-PST:TR=1 come-FUT:ITR
'I will beat up Jom and come back'
(N Zide 1997: 316)

The attempt to see how conjunctive participles are treated in these languages if they are present, did not meet much success. Most languages do not seem to have them and **Bhumij**, which has them, uses them for the stringing together of many clauses in a sentence, as is expected. While Ramaswami has translated them as finite subordinate clauses, Bhumij speakers use them as non-finite clauses with the verb showing no agreement or having no definitizer. **Santali** is slightly different in this respect. When two actions are reported as consecutive in nature, then the

two verbs representing the actions are ordered in their sequence of occurrence next to each other, without any conjunctive markers, and then the personal markers, tense/aspect, and definitizer, follow as though the two share these. This is clearly a phrasal attachment and therefore a characteristic clitic.

(22) *ɲel-ɲam=ked=a=e*
see-find=PST=DEF=3SG
"He looked/saw and found'
(Ghosh 1994:101)

3.1.4 Evidence from Serial Verb Constructions

In a study of clitics, serial verb constructions would be very informative on the status of the bound elements coming with these verbs. With serial verbs, the scope of certain affixes becomes a very pertinent issue. For example, if only one of the two serialized verbs carries tense marking but the tense applies to both the verbs, then with the scope of that marking extending to both the verbs it would be quite untenable to maintain that the marker is just an affix limited to the verb that hosts it. Considering the scope of the marker beyond the word that hosts it, it has to be treated as a phrasal affix, that is, as a clitic. In this regard, we look at AGREEMENT, TAM, DEF, SS, and DS markers, among others, to analyze their status as inflectional affixes or as clitics. Munda languages abound in such markers, which have to be rightfully analyzed as phrasal affixes due to their scope properties.

In **Mundari** serial verb constructions, the first lexical verb bears the subject enclitic (as a word immediately preceding the next verb) and the second verb comes with the DEFINITISER/FINITISER. The TENSE marker appears on the first verb. Since both FINITENESS and TENSE have semantic and syntactic scope over both the verbs, as both the verbs are to be read as finite as well as tensed, we have to treat these as phrasal affixes, that is, as clitics.

(23) *ne-nel=te=ŋ* *sen=a*
RDPL-see=T/A=1 go=DEF
'I will go and see'
(Hoffman 1903: 183)

Neukom (2001: 176), reports the following interesting example with respect to subject and object incorporation in serial verb constructions in Santali.

(24)

bʰəgtɛ=ko	*raɽa-led-e*	*ɲam-led-e*
quickly=3PL.SUBJ	release-PLUP:A-3SG.OBJ	find-PLUP:A-3SG.SUBJ
uni	*tərup-dɔ-e*	*rɔɽ-gɔt-ked=a*
that	tiger-TOP-3SG.OBJ	speak-V2-PST:A=DEF

'No sooner had they let him out and found him than the leopard/tiger said'
(Neukom 2001: 176),

In the above sentence, while the object agreement marker repeats itself on all the verbs, the subject agreement clitic appears just once,-on the word preceding the first verb- and the definitizer clitic appears also just once, but on the last verb of the series. We can also see that multiple verbs host the PLUP:A marker. That the object agreement marker appears repeatedly on all verbs, and so does the PLUP:A marker, with which it has a relationship, shows that they are typical inflectional affixes. In contrast, the appearance of the DEF marker only on the last verb but with a scope extending to all the verbs makes it a classic case of a phrasal affix, that is, a clitic.

In **Asuri (**Jashpur**)**, in serial verb constructions, the subject agreement marker appears on all the verbs in the series. The subject agreement marker behaves like a typical inflectional affix, with its scope relations met locally. However, the use of the DEF only once but with a scope on both the verbs makes it a fit case to be considered as a phrasal affix, that is, a clitic. However, since the agreement marker in (26) can be furthered from its host, by a clitic (=DEF), it has to be treated as a clitic, in a long distance relationship with its host.

(25)

sen-e-n=a:
go-ASP-TR=DEF
'He went'
(Grierson 1906: 139)

(26)

holate	*iŋ*	*huɽu*	*ir=iŋ*	*sen-tehin-en=a=iŋ*
yesterday	I	paddy	cut=1	go-T/A-ITR=DEF=1

'Yesterday I went and cut rice'
(Grierson 1906: 142)

In the following example of verb serialization in **Sora**, it is curious to find the agreement marker repeated both in the beginning and at the end of the verbal complex.

(27) *bagun-ben ə-il-le-ga-sal-n-e*
both-2PL 1/2PL-go-PST-drink-liquor-ITR-1PL
'Both of you went and drank liquor'
(Anderson and Harrison 2008: 360)

However, in the following examples from Sora, we see that the agreement and TAM markers appear only once, but with scope over all the verbs in the construction, giving them a phrasal affix (clitic) status. The double-marking of agreement could be a possible instance of the language moving its agreement markers from a clitic to an affixal status, as is also observed in some other Munda languages which have such alliterative markings.

(28) *paŋ-ti-dar=iɲ=teːn*
bring-give-cooked.rice=1=3.PST
'He brought and gave me cooked rice'
(Ramamurti 1931: 43)

(29) *anin iɲai=te=n-gu=am*
he come=NPST=ITR-call=2
'He came and called you'
(Ramamurti 1931: 44)

In Gtaʔ serial verb constructions or in constructions with complex verb stems, the agreement is seen to come only once. This will be difficult to explain unless; we give the marker the status of a phrasal affix with its scope on all the predicates – qualifying it to be a clitic, as in the following. In the following example, this can be contrasted to the RECIP marker, which appears on all the verbs where it is relevant and thus is a typical inflectional affix with its scope satisfied locally.

(30)

wiŋhaʔ=har=ke	*ho-m-m-og*	*ho-ʈ-m-u*	*ho-s-m-iʔ+ho-s-m-aʔ-har-ke*
quarrel=PL=T/A	RECIP-beat/ RECIP/	RECIP-throw.stone/ RECIP/	RECIP-cut/RECIP/+RECIP-catch/RECIP/-PL-T/A

'They beat each other, threw stones at each other, caught and butchered each other.'
(Anderson 2008: 720 [Mahapatra and Zide, no date, F. 17])

Gtaʔ does not have a very productive PERSON agreement marking system though it does mark NUMBER. However, it is important to note in the above example, the appearance of the NUMBER agreement marking and the T/A marking only once in a complex verb series with scope over all the verbs. The NUMBER agreement marking and the T/A marking, thus have to be treated as phrasal affixes (clitics).

In Gtaʔ, in a serial verb construction, all verbs except the last one have the SAME SUBJECT or DIFFERENT SUBJECT markers – the last verb carries the finite inflection; – it is attached to the verb like in other South Munda languages. The same subject (SS) marker in Gtaʔ may be followed by the enclitic [=*ka*] meaning 'only' or EMPHATIC (also found in Remo). The appearance of the TENSE marking only on one verb but with scope over all the other verbs in the following sentences must be taken to indicate a phrasal affix/clitic status for the TENSE marker.

(31)

ɖukri	*hoʔ-ru=hoʔ-ria=ce*	*swa*	*e-rro-raŋ=ce*
old.woman	weep=ECHO=SS	fire	go-RDPL:carry-bring=SS

hanɖa-nɖœ-ne	*moɽ-ke*	*cwar=ce*	*aʔ-nswar-bo=ke*
husband-3.REF-GEN	corpse-OBLQ	dry=SS	CAUS-dry-keep=ke.PST

'The old woman wept a lot and then made a fire, dried up her husband's corpse and preserved it' Anderson (2008: 750 [Mahapatra and Zide n.d.])

(32)

wig=la	*hɽiŋ*	*hanɖa-ŋɖe*	*pag=liʔ*	*we=ke*
go=DS	afterwards	husband-RFLXV	break=shoots	go=ke.PST

'She went and afterwards the husband went for bamboo shoots'
Anderson (2008: 753 [Mahapatra and Zide n.d.: 47])

(33) *hliʔ* *pag=ce* *coŋke=la* *poga* *sgwa* *bsœʔ* *lœʔ=ke*
shoot break=SS taste=DS tobacco like bitter AUX=ke.PST
'He broke the shoots and tasted them, they were bitter like tobacco'
Anderson (2008: 754 [Mahapatra and Zide n.d.: 47])

It is interesting that in Gtaʔ we do not find any productive agreement marking. However, the SS and the DS markers are used extensively but not/never on the final verb with which the actual scope of the SS and the DS markers rest. The DS marker is used when a verb has a different subject from that last verb; the SS marker is used when a verb shares the same subject as that last verb. But very curiously, most of the times, as the data seem to suggest, this last verb carries no marking of any agreement with the subject. Are we to assume a zero-marked verb? How do we interpret the SS and the DS markers, otherwise? The SS and the DS markers have scope over a larger constituent than the verb on which they appear and therefore are clear instances of phrasal affixes/clitics.

In multi-verb constructions in **Gutob**, either compounded or serialized, it is possible to have the agreement clitic on each verb or only on the final verb in the series, as can be seen in the examples below. With not all verbs carrying them, but their scope extending over all of them, these are phrasally attached clitics.

(34) *simra-gu* *ɖu-loŋ=nen*
enjoy-PST.I AUX-FUT.I=PL
'They will have enjoyed it'
(N Zide 1997: 314)

(35) *nom* *bobri-oʔ* *beɽ-oʔ*
you fill/CAUS/-PST.II AUX-PST.II
'You filled (it) up'
(N Zide 1997: 314)

(36) *loʈei-oʔ=niŋ* *beɽ-oʔ=su*
have.sex.w/woman-PST.II=1 AUX-PST.II=SS
'After I had sex with her…'
(N Zide 1997: 315)

(37) *loʈei=niŋ* *beʈ-oʔ=su*
have.sex.w/woman=1 AUX-PST.II=SS
'After I had sex with her…'
(N Zide 1997: 316)

(38) *tirgig=nei* *ɟuɟu=nei* *pi-loŋ-kina*
follow=1PL RDPL:see=1PL come-FUT.I-or.not
'Shall we come/follow along and see, come?'
(N Zide 1997: 310)

(39) *an-oʔ-su* *moʔʈ-gu* *piŋ-gi=niŋ*
pull.out-PST.II-SS get.up-PST.I come.back-PST.I=1
'I will pull it out, get up and come home'
(N Zide 1997: 316)

3.1.5 Evidence from Case Marking

All case markers, when bound, have to be seen as having scope over an entire noun phrase, and not merely over the word that it chooses as its host. Therefore, all bound case markers in the Munda languages have been treated as clitics in this study.

Mundari nominals/ pronominals bear no case marking. Their position within the VP/verb complex, or relative order outside the VP, gives an indication of their grammatical roles. The prototypical order of overt nominals is SUBJECT-INDIRECT OBJECT-DIRECT OBJECT, as can be seen in the following examples:

(40) *gomke* *dasi=e* *rak-i=a*
master servant=3SG call-3SG=DEF
'The master called the servant'
(Hoffmann 1903: 130)

(41) *gomke* *dasi* *talab=e* *oma-i=a*
master servant wage=3SG give-3SG=DEF
'The master gave the servant his wages'
(Hoffmann 1903: 130)

The other oblique case relations in Mundari are expressed with the help of bound post-positional markers, which owing to their scope over the entire phrase can be considered phrasal affixes and therefore treated as clitics.

In **Santali** too (cf: Ghosh (1994)), the nominative and the accusative are left unmarked with no overt case marking. There are overt post-positional or bound case markers for the INSTRUMENTAL, DATIVE, ABLATIVE, GENITIVE, and LOCATIVE. These bound forms, owing to their scope over the entire phrase, are considered phrasal affixes and therefore treated as clitics. Bound markers that appear after such phrasal affixes (like the INTENS marker in the example below) are also to be treated as clitics, as a bound morph coming after clitics cannot be considered integral enough to the structure of the word to be an affix.

The following examples demonstrate INSTRUMENTAL and DATIVE marking:

(42) *iɲ=tɛ=ge=ɲ* *mak-akan=a*
1SG=INSTR=INTENS=1SG cut-PERF=DEF
'I am cut by myself'
(Ghosh 1994: 42)

(43) *gɔrɔ=a=ɲ-mɛ*
help=DAT=1SG-2SG.IMP
'Help me'
(Ghosh 1994: 42)

It is important to note that the DATIVE marker [*=a*], in the example above, is not a real case marker – it can never form an independent word with the nominal/pronominal argument it marks. It satisfies more of a role that personal clitics play in the language – marking the presence of an argument in a particular syntactic/semantic role (indirect object/beneficiary). When used, it appears attached before the personal agreement clitics in the verbal complex. However, as a DATIVE marker, it has less to do with the verb that hosts it than to the pronominal marker in the verbal complex. The dative marked form is an already complex and derived form (*a=ɲ*), though not independent. This derived form is attached to the verb complex. As bound forms that are derived independently before attaching to the verbal complex, these bound markers at least together have some form of an independence from the verb, and therefore are to be treated as clitics. The DATIVE marked forms, when present, replace the ACCUSATIVE marked form and appear in that slot. See examples below:

(44) *gɔɽɔ=a=ɲ=mɛ*
help=DAT=1SG=2SG
'Help me!'
Ghosh (1994: 42)

(45) *iɲ* *ul=iɲ* *ɛm=a=m=a*
1SG mango=1SG give=DAT=2SG=DEF
'I will give you a mango'
Ghosh (1994: 61)

(46) *uni* *ləi=a=ko=a=e*
3SG tell=DAT=3PL=DEF=3SG
'He will tell them'
Ghosh (1994: 61)

In the above examples, since the dative marked forms are treated as clitics, bound markers that appear after such markers (like the DEF marker) are also to be treated as clitics, since clitics as post-inflectional markers are to be understood to have appeared after the end of the derivation of the word. This leaves no place for another bound marker to appear as an affix on the same word after a clitic. It should also be noted that in the above examples, there are no TENSE/ASPECT markers. When there are TENSE/ASPECT markers, according to Ghosh (1994: 61), the DATIVE marker fuses with them. And, as bound markers that appear after clitics, they are also to be treated as clitics, as one cannot have affixes coming after clitics.

(47) *uni* *ləi=a=d=iɲ=a=e*
3SG tell=DAT=PST=1SG=DEF=
3SG
'He told me'
(Here *a-d = a-ked*)
(Ghosh 1994: 61)

(48) *alaŋ* *hɔɽ=laŋ* *əgu=a=d=e=a*
2PL.INCL man=2PL.INCL bring=DAT=PST=3SG=
DEF
'We brought the man something'
(Ghosh 1994: 61)

(49) *uni alɛ sim=e kol-akaw=a=t=lɛ=a*
3SG 1PL cock=3SG send=DAT=PERF/PST=1PL=DEF
'He has sent us a cock'
(Ghosh 1994: 61)

In the above examples, the properties of personal clitics remain the same as in sentences without DATIVE arguments. The sentences above show all the arguments in their overt forms before the verbal complex, the last of which may take the subject marking enclitic, as in (48) and (49). The order of occurrence is: SUBJECT-IO-DO. Inside the verbal complex, the DO is not represented, following the general rule that if both ACC and DAT can be represented (because they are both animate), then by preference it is the DAT that gets represented inside the verbal complex.

A very important aspect that needs to be noted in the DATIVE constructions is that the DATIVE marker pushes the OBJ-DEF-SUBJ complex further away from the verb root, by coming in between the verb root and the TENSE/ASPECT marker. It would have been interesting if Santali had DATIVE subjects as well, like the Indo-Aryan and Dravidian languages of India, in which case it would have been interesting to know what form the SUBJECT CLITIC takes. In constructions, which have DATIVE subjects in Indo-Aryan and Dravidian languages, Santali seems to follow the syntax of normal assertive sentences, as in the following sentence, though it is not clear what the *li/l* attached to the 1SG is:

(50) *ruak-kan=a=liɲ*
fever-PROG=DEF=1SG
'I am getting fever'
(Ghosh 1994: 52)

In Ho, overt case marking is reported to be absent for the NOMINATIVE, ACCUSATIVE, and ERGATIVE functions. However, the BENEFACTIVE, INSTRUMENTAL, LOCATIVE, COMITATIVE, GOAL roles are reported to be expressed "by a case or clitic postposition following a noun or pronoun" (Anderson, Osada, and Harrison: 2008: 206). Though the authors do not say why they think they could be clitics, one of the arguments that can be used to argue a clitic status for some of these markers is the optionality in their use, as can be seen in the following example of an instrumental marking (also used for ablative marking). The more important reason though is the phrasal nature of their attachment to their hosts:

(51) *jilike=m* *hujuʔ-ye-n-e*
how=2 come-T/A-ITR-DEF
'How did you come (here)?'
(Anderson, Osada and Harrison 2008: 207)

(52) *jilike=te=m* *hujuʔ-ye-n-e*
how=INSTR/ABL=2 come-T/A-ITR-DEF
'How did you come (here)?'
(Anderson, Osada, and Harrison 2008: 207)

Not all Munda languages report an absence of overt marking of nouns/pronouns in the subject or object positions. Case is marked in **Gorum**, though only on object forms of pronominals, with a prefixal/proclitic*e*=, as can be seen in such pairs as *nɔʾd-enɔʾd* [3SG.NOM-3SG.ACC], *mɔy-emɔy* [who-to whom] (Anderson and Rau 2008: 391-92). Such bound case forms, due to their scope extending over an entire phrase, are considered clitics. Case is also marked post-positionally, using *etur, ɔtur*, etc., to signify locational/genitive relations (Anderson and Rau 2008: 388). However, it is also noted that markers like *etur* are found obligatorily only with nominals while they appear optionally with pronouns and inanimate nouns (Anderson and Rau 2008: 389). If *etur* is an independent word and not a bound form, as is represented in our source, such optionality will be insignificant for us. There is as such no evidence of a clitic status for any of the overt case markers found in Gorum. However, this is only true as long as we can accept the analysis of these as given in the texts. However, the function or placement of a marker like *etur* is far from clear, if one were to look at the sample sentences that appear below (Anderson and Rau 2008: 424-25):

(53) *zɔtɔn* *ɖa-ru* *ɖu* *rusi-ɖi* *babey* *ɔ*
guardian do-PST and priest-FOC think.AFF that

kuntur *e-nɔʔd* *etur* *kua-nen* *zum-tu* *laʔ-tu*
rat OBJ-s/he OBJ crow-this eat-NPST AUX-NPST

'Having taken care of him, the priest thought "that shrew, this crow will eat him up, for sure"'

(54) *miŋ* *bɔʔptoŋ-t-ay* *bɔyragi-ɖi* *etur* *sun-ru* *ɖruka-ɖi*
1SG frighten-NPST-CLOC priest-FOC OBJ say-PST tiger-FOC

'"I will frighten you" said the tiger to the priest.'

(55) *dinek* *gɔrzɔn* *ɖa-ru* *ɖu* *boyragi-ɖi* *etur*
one.day roar do-PST and priest-FOC OBJ
leʔn-u *zum-u* *ɖaʔd* *ui*
catch-INF eat-INF for go.AFF
One day he roared and went to catch and eat the priest'

In **Sora** too, as the following examples demonstrate OBJECTIVE case is marked as a prefix on all nouns and pronouns, and because of its potential scope over an entire phrase is treated as a clitic:

(56) *kuni* *a=tarbaɲ-ɟi* *a=maɳɖra* *tij-a*
those OBJ=flower-PL OBJ=man give-IMP
'Give those flowers to the man'
(Bhattacharya 1975: 162)

(57) *bab-ɲen* *a=uʔ*
head-1 POSS=hair
'the hair of my head'
(Bhattacharya 1975: 169)

In his description of **Kharia** morphology, Peterson (2008: 441) considers many bound forms in the language to be clitics/enclitics, including case markers. He draws our attention to the fact that the schematic overview of the NP in Kharia would be as follows:

GEN DET DEM QUANT GEN DET LEXEME(s)= POSS= NUM= CASE

One can notice that the elements listed on the right end are mostly clitics. Peterson justifies this classification/nomenclature in the following words: "… if the lexical base is not overtly present (e.g. if it is unknown), these markers simply attach to the right-most element of the phrase, regardless of its status, whether lexical or genitive attribute. That is, unlike affixes, these markers do not attach to stems. Rather, they attach to the entire phrase" (441). The above description can easily be taken as one of the most defining characteristics of clitics.

The following examples show the status of these bound markers as clitics:

(58) *aɖi=yaʔ* *gʰol* *beʈa=ɖom=ki=te*
ANAPH=GEN ten son=3POSS=PL=OBLQ
'His ten sons'
(Peterson 2008: 441)

(59) *aḍi=yaʔ* *moŋ* *dʰaŋgar=ḍom=te* *gam=oʔ*
ANAPH=GEN one servant=3POSS=OBLQ say=A.PST
'He said to one of his servants'
(Malhotra 1982: 127)

Peterson also notes that if no further information exists within the NP, on account of them being known from the context "these markers attach directly to the demonstratives, which then serve as pronominals of the third person, for example, *ho=ki=te* (that=PL=OBLQ) 'them', from 'that'" (441). The movement and attachment of clitics to the right-most ending phrase in situations where the host word is dropped because it can be retrieved from the context of the discourse can be observed with number marking as well. Here, we will focus on the use of case markers and their status as possible clitics.

Peterson considers the overtly marked oblique and genitive cases to be enclitics. The oblique marker *=te*is seen to be marking definite objects, indirect objects, and adverbials, and the genitive is marked by *=(y)aʔ*. While the guiding principle in declaring number markers as clitics is an ability to attach to elements unrelated to them, when their hosts are dropped in familiar contexts, the case markers are argued to "attach directly to the bare nominal (or rather, the last element of the lexical base of the NP)" (443). Peterson also notes that both, the oblique and genitive case marking clitics, can appear together in one particular type of construction. In this construction "if the semantic head of the NP is not overtly expressed, the (enclitic) oblique marker *=te* attaches to the right-most element of the remaining lexical base of the NP regardless of its status. If this element is a genitive determiner, this results in apparent 'double case marking'" (443), as in:

(60) *iŋ=aʔ=te* *saykal* *ayiʔj*
1SG=GEN=OBLQ bicycle PRS.COP
'I have a bicycle'
(Peterson 2008: 443)

The structure above is a reduced form of the structure below:

(61) *iŋ=aʔ* *boʔ=te* *saykal* *ayiʔj*
1SG=GEN place=OBLQ bicycle PRS.COP
'I have a bicycle' (lit. There is a bicycle at my place)
(Peterson 2008: 443)

So, one can see, that the basis on which number and case markers are considered enclitics are similar (except that, bound case markers are considered clitics primarily because they are phrasal affixes), and there seems to be a definite pattern that these elements follow, which arguably makes them fit cases to be considered clitics.

The bound nature of the **Juang** case markers, like the GENITIVE and the DEFINITENESS marker, make them clitics. This is because; case markers have a scope that goes beyond the hosting word.

(62) *aiɲ=a* *uŋi=n=ḍe* *kɔsɔkɔ=ra* *ele* *ape=a* *sɔpa=rɔ*
1SG=GEN shirt=1=DEF dirty=DEF but 2PL=GEN clean=DEF
'My shirt is dirty, but yours is clean.'
(Patnaik 2008: 519)

Juang subjects, when in the DATIVE, show no agreement with the verb (Patnaik 2008: 519) as can be seen in the following example.

(63) *araki* *araka+araka* *diɔ* *ku-buji=ri=ki*
3PL each.other well RECIP-do=DEF=PL
'They love each other.'
(Patnaik 2008: 523)

If we look at CASE marking in **Remo**, it is noted to have an objective case marker *a=* which is considered a very unusual feature in Remo nominal morphology by scholars. Anderson and Harrison (2008) note that this case marker "is nearly obligatory with pronouns and in a number of contexts with nouns as well" (570). More importantly, they note that "structurally speaking, the *a*-OBJECTIVE is not a prefix, but rather a proclitic which targets the leftmost edge of the relevant NP that it seems to mark. Thus, it may appear on a demonstrative pronoun, a possessive pronoun, or anything likely to precede a noun or come initially in a Remo noun phrase" (572). They are thus treated as clitics for their scope properties. Following are examples of the use of this marker:

(64) *a=kon* *bire* *kur*
OBJ=that stone roll
'Roll down that stone'
(Fernandez 1968: 67)

(65) *a=kon soka oyja dabu ɖi-ta*
OBJ=that shirt how.much money COP-NPST.II
'How much does that shirt cost?'
(Fernandez 1968: 67)

(66) *niŋ a=niŋ-ŋa ɖio uriŋ-t-iŋ*
1 OBJ=1-GEN house walk-NPST-1
'I will walk to my house.'
(Fernandez 1968: 119)

(67) *gitin remo a=monaʔbay selane kiyaŋ beɖ-oʔ*
that.CLOSE man OBJ=fat girl rice give-PST.I
'That man gave rice to the fat girl.'
(Fernandez 1968: 119)

Anderson and Harrison (2008) also refer to the adpositions in Remo as bound/enclitic postpositional or case elements. Though the authors do not make it clear, it is being assumed that the clitic status is probably based on the phrasal nature of the affixation of the bound elements. Postpositions or prepositions, as heads of adpositional phrases, have scope over the entire phrase.

The clitic hood of case markers in Gtaʔ needs to be understood in terms of its position in a word/phrase and its interaction with number marking. Number markers usually come attached to the noun stem. However, the number marking can be separated from the stems when inalienable possession is marked. This is difficult to explain unless we assume a phrasal affix-like nature for the plural marker, which makes it a clitic. And, if the case marker appears in the noun phrase after such a clitic-like marker, it also has to be understood as phrasal/enclitic. The following example (where the PLURAL or the CASE marker is not treated as clitics in the original source), shows the case marker coming after the number marking enclitic:

(68) *huŋ-ɖœ=hiŋ=ke*
child-3=PL=CASE
'(to) his children'
(Anderson 2008: 688)

Gtaʔ has no object marking, except for the 3PL marker *=har*, marking the plurality of objects in some constructions, as in the following sentence. As a bound marker, with its scope extending over a phrase, it is treated as

a clitic. The tense marker following a clitic in the example below has also to be treated as a clitic.

(69) *gte-la nœŋ mriaʔ-ce a=mœ-hiŋ-ke m-bagweʔ =har=e*
then I rise-SS OBJ=3-PL-CASE 1-kill=PL:OBJ=FUT
'Then, I will get up and kill them all.'
(Anderson 2008: 723 [Mahapatra and Zide, no date, D. 11])

The status of the OBJECTIVE marker *a=* in Gtaʔ in the above structure needs to be considered in consonance with the analysis of the same in Gutob and Remo as a proclitic, for the reasons advanced for them there.

The OBJECTIVE marker *o=* in **Gutob** has been considered a proclitic for the same reasons as the parallel marker *a=*in Remo and the OBJ marker *a=* in Gtaʔ have been considered proclitics by Anderson and Harrison (2008) and Anderson (2008) respectively. The nature of affixation being phrasal, these are rightly treated as clitics.

3.1.6 Miscellaneous Clitics

In (70) below, the Q particle *cʰi* in Mundari is considered an enclitic because its scope extends over the entire clause, and therefore, it is not a word-level affix but a phrasal affix.

(70) *sena=m=cʰi*
go=2SG=Q
'Will you go?'
(Hoffmann 1903: 71)

It can be noticed that the general pattern for the enclitic in Bhumij is to attach to the word preceding the verbal complex. Sentence (71) below is a compound sentence and the agreement enclitic is seen to repeat for each of the predicates – in the first instance appearing on the verb itself and in the second appearing on the word preceding the verb. (71) is also a very important structure as far as the clitic status of the DEF marker is concerned. The DEF/COP appearing only once is an instance of its scope over all the elements of a coordinated structure, making it clearly a phrasal attachment and, therefore, a clitic. This is as per the criterion discussed in Miller (1992) with respect to recognizing clitics in coordinated structures.

(71) *aiŋ sen=iŋ ar=iŋ auy=a*
1SG go=1SG CONJ=1SG bring=COP
'I will go and bring'
(Ramaswami 1992: 131)

Clear evidence of the agreement markers in Gutob being clitics comes from the marking of PROH NEG. The PROH NEG in Gutob is not a separate word but an add-on to regular negation. As a bound particle within the verbal complex, with a scope over the entire clause, the PROHIBITIVE NEGATION marker is a clitic. It also, by appearing next to the verb stem, furthers the subject agreement marker from the verb, as in the following sentence. As a bound inflectional marking appearing immediately next to a clitic, which is a phrasal affix by virtue of its scope properties, the agreement marker even though bound has to be seen as being affixed phrasally. Therefore, it is clearly a phrasal affix and therefore a clitic.

(72) *a=ḍo=ge=pe*
NEG=flee=PROH=2PL
'Don't flee!'
(Anderson 2008: 706 [Mahapatra and Zide, no date, J. 23])

4.0 Conclusion

Due to the limitations of space, a large number of phrasal affixes in the Munda languages that can be analysed as clitics have not been presented here. However, it is hoped that this brief discussion on the possibility of recognizing a wide array of particles as clitics because of their morpho-syntactic properties and their phrasal level of attachment, opens up an important dimension to the study of clitics in not only the Munda languages but also in other highly agglutinating languages.

Notes

[1]Available at <http://www-01.sil.org/linguistics/glossaryoflinguisticterms/whatisacliticgrammar.htm>

Abbreviations

-	AFFIX BOUNDARY	**DEM**	DEMONSTRATIVE
.I	FIRST CONJUGATION	**DS**	DIFFERENT SUBJECT MARKER

.II	SECOND CONJUGATION
1PL	FIRST PERSON PLURAL PRONOMINAL OR AGREEMENT CLITIC
1SG	FIRST PERSON SINGULAR PRONOMINAL OR AGREEMENT CLITIC
2PL	SECOND PERSON PLURAL PRONOMINAL OR AGREEMENT CLITIC
3PL	THIRD PERSON PLURAL PRONOMINAL OR AGREEMENT CLITIC
A.PST	ACTIVE PAST
ABL	ABLATIVE CASE MARKER
ACC	ACCUSTAIVE CASE MARKER
ANAPH	ANAPHORIC PRONOMINAL
AOR	AORIST
ASP	ASPECTUAL MARKER
AUX	AUXILIARY (VERB)
BEN	BENEFACTIVE
CASE	CASE MARKER
CAUS	CAUSATIVE MARKER
CLOC	CISLOCATIVE MARKER
CONJ	CONJUNCTION
COP	COPULA
DAT	DATIVE CASE MARKER
DEF	DEFINITIZER
PST	PAST
Q	QUESTION PARTICLE
QUOT	QUOTATIVE
RDPL	REDUPLICATED FORM
RECIP	RECIPROCAL MARKER
RFLXV	REFLEXIVE MARKER
FOC	FOCUS PARTICLE
FUT	FUTURE TENSE
GEN	GENITIVE CASE MARKER
IMP	IMPERATIVE
INCL	INCLUSIVE
INF	INFINITIVE
INSTR	INSTRUMENTAL
INTENS	INTENSIFIER
ITR	INTRANSITIVE
LOC	LOCATIVE CASE MARKER
NEG	NEGATIVE PARTICLE
NF	NON-FINITE
NOM	NOMINATIVE
NPST	NON-PAST
OBJ	OBJECT
OBLQ	OBLIQUE CASE MARKER
PERF	PERFECT ASPECT MARKER
PL	PLURAL NUMBER
PLUP	PLUPERFECT
POSS	POSSESSIVE MARKER
PROG	PROGRESSIVE ASPECT MARKER
PROH	PROHIBITIVE
PROH:NEG	PROHIBITORY NEGATION
PRS	PRESENT
SS	SAME SUBJECT MARKER
SUBJ	SUBJECT AGREEMENT MARKER
T/A	TENSE/ASPECT
TOP	TOPICAL MARKER
TR	TRANISITIVE MARKER

References

Anderson, Gregory, D.S. *The Munda Verb. Typological Perspectives*. Berlin: Mouton de Gruyter (Trends in Linguistics, Studies, and Monographs, 174), 2007.

—. "Gta?." In *The Munda Languages*, edited by Gregory D.S. Anderson. 682-763.London & New York: Routledge, 2008.

—, and Felix Rau. "Gorum." In *The Munda Languages*, edited by Gregory Anderson D.S., 381-433. London & New York: Routledge, 2008.

—, and K. David Harrison. a. "Remo (Bonda)." In *The Munda Languages*, edited by Gregory Anderson D.S., 557-632. London & New York: Routledge, 2008.

—, and K. David Harrison. b. "Sora." In *The Munda Languages*, edited by Gregory Anderson D.S., 299-380. London & New York: Routledge, 2008.

—, Toshiki Osada, and David Harrison. 2008. "Ho and other Kherwarian languages." In *The Munda Languages*, edited by Gregory Anderson D.S., 195-255.London & New York: Routledge, 2008.

Anderson, Stephen R. *A-Morphous morphology*. Cambridge: CUP, 1992.

—. "Wackernagel's revenge: Clitics, morphology and the syntax of second position". *Language* 69 (1993): 68-98.

—. *Aspects of the theory of clitics*. Oxford: Oxford University Press, 2005

Bhattacharya, Sudhibhushan. "A new classification of Munda." *Indo-Iranian Journal* 17(1975): 97-101.

Fernandez, Frank. "A grammar of Remo." Ph.D. thesis, University of North Carolina, 1968.

Ghosh, Arun Kumar. *Santali -A Look into Santal Morphology*. New Delhi: Gyan Publishing House, 1994.

Grierson, G.A. *Linguistic Survey of India*, Vol. 4, Munda, and Dravidian. New Delhi: Low Price Publications, 1906.

Halpern, Aaron L. *On the placement and morphology of clitics*. Stanford: CSLI Publications, 1995.

Hoffmann, John. *Mundari Grammar*. Calcutta: Bengal Secretariat Press, 1903.

Klavans, Judith L. "The Morphology of Cliticization." In *Papers from the parasession on the interplay of phonology, morphology, and syntax*, edited by John F Richardson, Mitchell Marks, and Amy Chukerman, 103-121. Chicago: Chicago Linguistic Society, 1983.

Klavans, Judith L. "The independence of syntax and phonology in cliticization." *Language* 61(1985): 95-120.

Malhotra, Veena. “The Structure of Kharia: A Study of Linguistic Typology and Language Change.” Ph.D. dissertation, Jawaharlal Nehru University, New Delhi, 1982.

Miller, Philip. *Clitics and constituents in phrase structure grammar* [Outstanding Dissertations in Linguistics]. New York: Garland, 1992.

Neukom, Lukas. *Santali*. München: Lincom Europa, 2001.

Osada, Toshiki. *A reference grammar of Mundari*. Tokyo: Tokyo University of Foreign Studies, Institute for the Study of Languages and Cultures of Asia and Africa, 1992.

Osada, Toshiki. “Mundari.” In *The Munda Languages*, edited by Gregory D.S. Anderson, 99-164., London & New York: Routledge, 2008.

Peterson, John. “Kharia”. In *The Munda Languages*, edited by Gregory D.S. Anderson, 434-507. London & New York: Routledge, 2008.

Ramamurti, G.V. *A manual of the So:ra (or Savara) language*. Madras: Government Press, 1931.

Ramaswami, N. *Bhumij Grammar*. Mysore: Central Institute of Indian languages, 1992.

Zide, Norman H. “Gutob pronominal clitics and related phenomenon elsewhere in Gutob-Remo-Gta?.” In *Languages of tribal and indigenous people of India: The ethnic space*, edited by Anvita Abbi, 307-334. New Delhi: MLBD, 1997.

Zwicky, Arnold M. *On clitics*. Bloomington: Indiana University Linguistics Club, 1977.

Zwicky, Arnold M. “Suppressing the Zs”. *Journal of Linguistics* 23(1987): 133–148.

CHAPTER EIGHT

PHONETIC COMPARISON OF ORISSA SORA AND ASSAM SORA

LUKE HORO AND PRIYANKOO SARMAH

1. Introduction

The Sora language is originally spoken in Orissa (in eastern India), and Assam Sora emerged due to the migration of some Sora groups from Orissa to Assam (in northeast India) as indentured tea labourers in the 19th century. Geographically, Orissa and Assam are over a thousand kilometers apart, and in the past hundred years, the Soras of Assam have nearly lost all plausible contact with their ancestors in Orissa. Also, the migrant Sora community in Assam is surrounded by speakers of different languages that are not present in Orissa. Diachronically, migration is an important extra-linguistic factor of language change (Kerswill: 2006). There is evidence that migrated languages, also known as transplanted languages, change and become distinct from the original language. In this regard, language contact is often the dominant factor causing language change in transplanted languages (Trudgill: 1994). Hence, considering that Assam Sora emerged from the migration of Sora groups from Orissa to Assam, this paper examines the impact of migration on Assam Sora to see if language contact has changed Assam Sora or not. For this purpose, the study compares three segmental properties of Sora as it is spoken in Orissa and Assam (to be referred to as Orissa Sora and Assam Sora in this paper) with the help of synchronic speech data and acoustic analysis. The segmental properties compared in this work include vowel inventory sizes, phonetic realisations of word stress, and voicing contrast of stop consonants. These properties are selected since they can adequately distinguish between languages as well as between language varieties. For instance, there is evidence that some languages can be divided into iambic and trochaic based on the placement of word stress in those languages

(Hayes: 1985). In the case of vowels, Williams and Escudero (2014) show that cross-dialectal variation between northern and southern British English can be explained by comparing their vowel trajectories. Likewise, Holliday and Kong (2011) show that word-initial stop consonants are produced differently by speakers of different Korean dialects. Hence, the current study explores these three properties in Sora to examine how the phonetic properties of Sora of Orissa are related to the Sora of Assam at this point.

2. Background

2.1 Vowels in Assam Sora and Orissa Sora

Our previous study (Horo and Sarmah:2015) of Assam Sora particularly the vowel data, reveals that Assam Sora has six vowel phonemes including /i, e, o, a, u, ə/. On the other hand, the literature on Orissa Sora has a diversity of vowel descriptions. Firstly, Stampe (1965), Zide (1982), and Donegan (1993) propose that Orissa Sora has nine vowel phonemes including /i, e, a, o, u, ə, ɛ, ɔ, ɨ/. In addition, studies such as Anderson and Harrison (2008a) identify eight vowel phonemes /i, e, a, o, u, ə, ʊ, ɨ/ in Orissa Sora and works such as Ramamurti (1931 and 1938) propose that in addition to the six vowel phonemes /i, e, o, a, u, ə/ Orissa Sora also has allophonic vowels such as /ü, ɪ, ö/ and that there is a three-way vowel length contrast for all vowels including the allophonic variations. Further, Mohanty (1997) speculates that Orissa Sora has only five vowels /i, e, a, ɔ, u/ and proposes that Orissa Sora vowel inventory is influenced by an areal typology whereby three languages namely, Kui (a Dravidian language), Sora (a Munda language) and Oriya (an Indo Aryan language) have developed uniform vowel inventories. Furthermore, in the typological overview of Austroasiatic languages presented by Jenny et al. (2014:30), it is stated that *Mundari, Kera?, Korku, Kharia, Sora and Gutob have a 5-vowel system /i, u, e, o, a/* and that, *the central unrounded non-low vowels /ə, ɨ/ are absent in most languages but Sora has both.* Thus, by surveying the relevant literature on Orissa Sora, it is clear that different scholars have proposed different vowel inventories of Orissa Sora. While consensus regarding a nine-vowel system of Orissa Sora proposed by Stampe (1965) and by Donegan (1981) is found in the works of Zide (1982) and Donegan (1993), Ramamurti (1931 and 1938) it is often criticized due to inadequate evidence. Other studies such as, the work of Anderson and Harrison (2008a) who indicate that there are eight vowel phonemes have not been discussed by other scholars so far. Also, the arguments of Mohanty (1997)

have not been reviewed so far. Hence, it is clear that although the vowel features of Orissa Sora have been described in the past, the descriptions are not consistent and it is difficult to make an adequate comparison between Assam Sora and Orissa Sora based on the available literature. Therefore, the current study makes a synchronic comparison between the two Sora varieties based on their present-day speech data.

2.2 Word Stress in Assam Sora and Orissa Sora

Analysis of word stress in Assam Sora (Horo and Sarmah: 2015) revealed that in disyllabic words the second syllable is always prominent and prominence of the second syllable is realised by greater vowel duration, greater fundamental frequency, and greater vowel intensity. Thus, it is evident that there is a dominant iambic stress pattern in Assam Sora disyllabic words. On the other hand, the stress pattern of Orissa Sora is undescribed. However, Donegan (1993) and Donegan and Stampe (2004) propose that Munda languages have trochaic stress patterns because the typology of South Asian languages influences them. Significantly, this generalisation is not found in all Munda languages. There is evidence that languages such as Santali (Ghosh 2008) and Plains Remo (Anderson and Harrison 2008b) have stress on the second syllable. Hence, by comparing the stress patterns of Orissa Sora and Assam Sora, this study can evaluate if Orissa Sora has a South Asian stress pattern namely trochaic and if Assam Sora developed an iambic stress pattern as a result of migration. Also, this examines whether or not Orissa Sora too has an iambic stress pattern similar to other Munda languages such as Santali and Plains Remo.

2.3 Stop Consonants in Assam Sora and Orissa Sora

All scholars agree that Orissa Sora has six stop consonants including /p, t, k, b, d, g/ that have two-way voicing contrast. Also, it is agreed that the stop consonants /t/ and /d/ are asymmetric in Orissa Sora such that while /t/ in Orissa Sora is a voiceless dental stop, /d/ in Orissa Sora is a voiced alveolar stop. Also, Stampe (1965) suggests that this asymmetry is, in fact, a core feature of Orissa Sora which accounts for it to be included in the Munda subgroup of the Austroasiatic language family. Moreover, Ramamurti (1938) suggests that the voicing asymmetry between /t/ and /d/ is often neutralised due to allophonic variations. Evidence suggests that the voiceless dental stop /t/ can change to a voiceless alveolar stop [t] in the environment of alveolar sounds, and the voiced alveolar stop /d/ can

change to a voiced dental stop [d̪] in the environment of dental sounds. Significantly, our preliminary observations suggest that Assam Sora also has the same six stop consonants. However, the existence of /t/ and /d/ voicing asymmetry does not seem to be consistent in Assam Sora. It appears that unlike Ramamurti (1938) suggests, the asymmetry is neutralized irrespective of environmental influences. This indicates that Assam Sora stop consonants are likely to differ from Orissa Sora stop consonants to a certain extent. Hence, the current study examines Assam Sora and Orissa Sora stop consonants in terms of their voicing contrasts in the word-initial position and confirms whether synchronically the two Sora varieties have the same /t/ and /d/ voicing asymmetry or not.

3. Methods

3.1 Data and Recording

The data set consists of 202 Sora words that include data for minimal vowel set, data for word-initial stop consonants, and data for different disyllabic words in Assam Sora and Orissa Sora. The vowel minimal set also includes all vowels in word-initial, medial, and final positions. Also, data for analysing word-initial stop consonants include all six oral stops preceding all six vowels in Assam Sora and Orissa Sora. Primary sources of the data set are Ramamurti (1938), Anderson and Harrison (2011), and Horo and Sarmah (2015). The entire word list was first verified with some Assam Sora and Orissa Sora participants, and then all the participants were recorded saying every word once in isolation and once in the sentence frame 'ɲen_____ gamlai'translated as *'I_____said'*. However, although the words are recorded in isolation as well as in a sentence frame, the analysis here considers only the samples that are recorded in isolation.

3.2 Location and Participants

In Orissa, the field study was conducted in the Rayagada district of southern Orissa that is adjacent to parts of Vizianagaram and Srikakulam district of Andhra Pradesh. Initially, a few interviews were conducted in a village named Marichaguda under Padmapur block but the speech data was recorded in a village named Raiguda under Jagannathpur Gaon Panchayat in Gunupur block of Rayagada district. Raiguda is a Sora-dominated village in the Rayagada district of Orissa. There are approximately 85 houses in the village. People in the village mostly speak Sora and Oriya, and some could also speak Telugu. The area was selected

in consultation with Assam Sora informants and one Assam Sora informant also participated in the field study as a Sora interpreter in Orissa. Five Sora males and five Sora females of Raiguda village were recorded while saying the dataset mentioned in 3.1.

In Assam, the field study was conducted in the Lamabari tea estate of Udalguri district and Koilamari tea estate of Lakhimpur district. Both Lamabari and Koilamari tea estates are Sora concentrated villages in the two districts of Assam and Sora groups in both the areas have reported that they have migrated to Assam from the undivided Ganjam district of Orissa in the 19th century. There are approximately 800 Sora speakers in the Lamabari tea estate and 300 Sora speakers in the Koilamari tea estate. The Sora speakers in both areas can speak Sadri in addition to Sora, and some Sora speakers can also speak Assamese and Hindi in addition to Sora and Sadri. Ten participants (five male and five female) were recorded for the same data set mentioned in 3.1 in Lamabari as well as in Koilamari.

3.3 Acoustic Measurements

Digital speech data recorded from the field studies are annotated in Praat for word boundaries, syllable boundaries, and phoneme boundaries. Phoneme boundaries for all vowel tokens are marked between the beginning and the end of steady-state vowel formants. Similarly, every word-initial stop consonant is marked between the release of the stop consonant and the onset of the glottal pulse of the following vowel. Subsequently, all acoustic analysis is done in Praat. Acoustic analysis of vowels is based on formant frequencies of the first two formants (F1 and F2), whereby formant frequencies are extracted from steady-state formants at the vowel mid-point. Word stress in disyllabic words is analysed by measuring the temporal properties of the syllable nuclei. For this purpose, average fundamental frequency (*f*0 or pitch variation), average vowel duration (the time interval), and average vowel intensity (the acoustic energy) are calculated between the beginning and end of glottalic pulses for all syllable nuclei in the first and second syllable of disyllabic words. Finally, analysis of word-initial stop consonants is based on the VOT (Voice Onset Time) values of the stop consonants whereby VOT captures the time interval between the release of an oral stop consonant and the onset of the glottal pulse of the following vowel.

4. Results

4.1 Vowels

In the background of this study, it is shown that researchers in the past had not reached a consensus regarding the vowel phoneme inventory of Orissa Sora. Scholars have proposed that there are five, six, eight, or nine vowel phonemes in Orissa Sora. In this regard, the examination of a vowel minimal set, showing nine contrastive vowels in Orissa Sora (Donegan and Stampe,2002), indicates that there is either misrepresentation of the vowel phonemes or that some of the vowel phonemes have assimilated to the adjacent vowels in Assam Sora (see Table 1).

Vowel	Orissa Sora (Donegan and Stampe 2002)	Assam Sora	English
/i/	[iɖ-]	[id-]	scratch
/ɨ/	[ɨɖ-]	[əʔ-]	fan
/ə/	[əɖ-]	[əd-]	prop
/e/	[-eɖ-]	[iʔ-]	thorn
/o/	[-loɖ-]	[lud]	cord
/ɛ/	[ɛɖ-]	[ed-]	roll
/ɔ/	[ɔɖ-]	[od-]	knead
/ɑ/	[ɑɖ-]	[ad-]	drive
/u/	[-luɖ-]	[luʔ]	ear

Table 8- 1: Comparing Orissa Sora vowel data with Assam Sora

It is evident from Table 1 that the majority of Orissa Sora vowels: /ɨ, ɛ, ɔ, e, o/ are produced differently in Assam Sora. Significantly, these alterations are also observed in the synchronic Orissa Sora speech data recorded in this study. Thus, it is apparent that there may have been misrepresentations of the vowel phonemes of Orissa Sora in the past that gave rise to a vowel inventory of nine vowels in Orissa Sora. Therefore, this study considers the vowel minimal set of Assam Sora (Horo and Sarmah, 2015), that proposed Assam Sora has six vowel phonemes /i, e, ə, o, u, a/, and compares the vowel phonemes of Assam Sora and Orissa Sora based on Assam Sora vowel data. During the field study, in Raiguda as well as in

Lamabari and Koilamari it has been observed that the same six vowels are produced distinctly by all the participants recorded in this study.

Further, acoustic analysis of the vowel data, reveals that every participant recorded in this study contrastively produces the vowels /i, e, ə, o, u, a/. Figures 1, 2, and 3 present the Lobanov normalised vowel plots of Orissa Sora and Assam Sora based on their F1 and F2 formant frequencies.

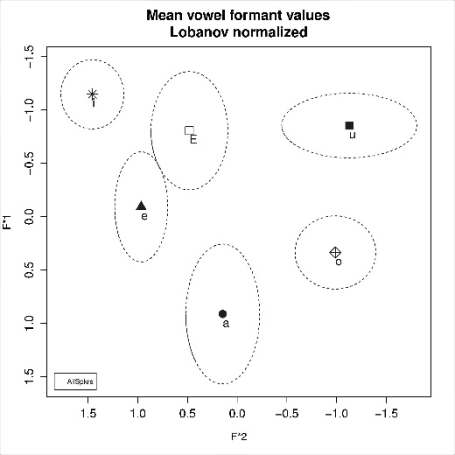

Figure 8-1: Assam Sora Vowel Plot (Koilamari)

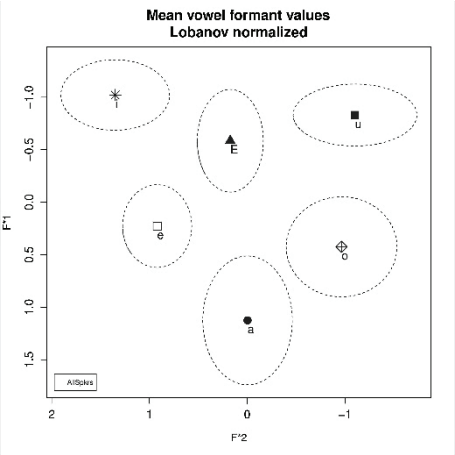

Figure 8-2: Assam Sora Vowel Plot (Lamabari)

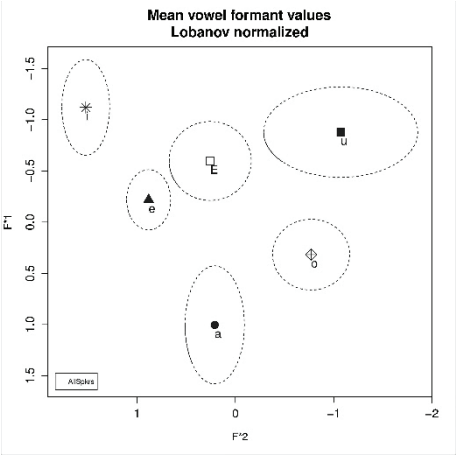

Figure 8-3: Orissa SoraVowel Plot (Raiguda)

It is evident from figures 1, 2, and 3that the vowels /i, e, ə, o, u, a/ are perceptually contrastive in Assam Sora as well as in Orissa Sora. Hence, this study proposes that similar to Assam Sora, Orissa Sora has six contrastive vowel phonemes. This study supports the claim with evidence drawn from the acoustic analysis of the vowel phonemes in Orissa Sora and Assam Sora.

4.2 Word Stress

Our previous analysis of Assam Sora disyllabic data revealed that the second syllable is always prominent (Horo and Sarmah, 2015). The pattern was evident from the higher fundamental frequency, longer vowel duration, and higher vowel intensity of the second syllable in comparison to the first syllable in disyllables. Therefore, disyllabic Sora data from Raiguda, Lamabari, and Koilamari were subjected to similar measurements to compare word stress in Orissa Sora and Assam Sora in this study. Analysis of disyllabic data from the three regions revealed that phonetic realisation of word stress in Orissa Sora is similar to Assam Sora. Firstly, Figures 4, 5, and 6 show that average vowel duration is always greater in the second syllable in Orissa Sora as well as in Assam Sora.

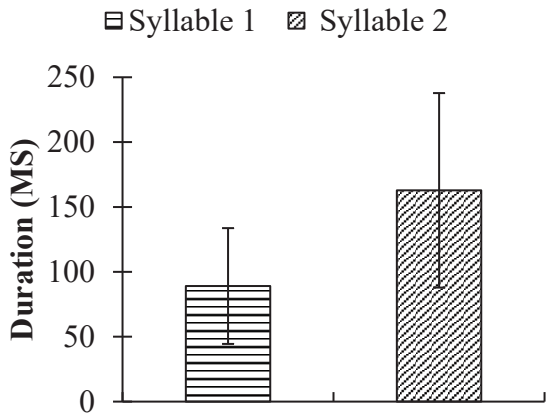

Figure 8-4: Average Vowel Duration Assam Sora (Koilamari)

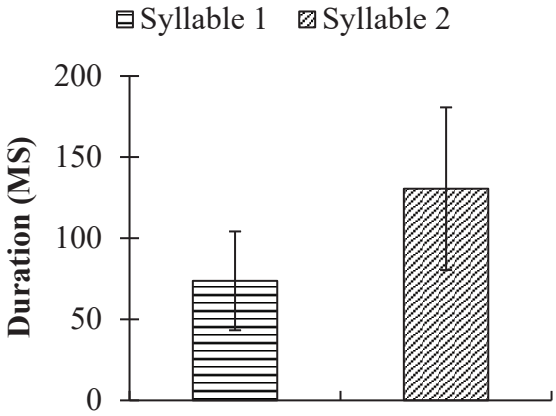

Figure 8- 5: Average Vowel Duration Assam Sora (Lamabari)

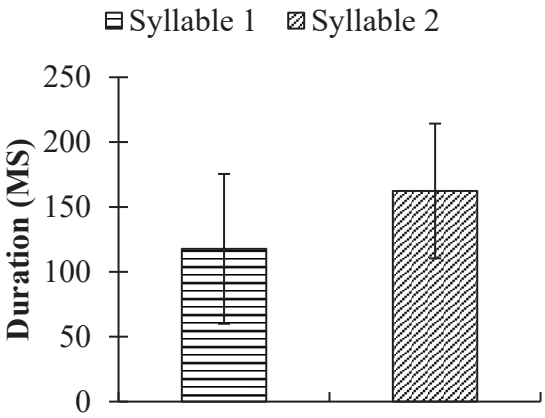

Figure 8-6: Average Vowel Duration Orissa Sora (Raiguda)

Similarly, Figures 7, 8, and 9 show that the average *f*0 of the second syllable is generally greater in Orissa Sora and Assam Sora except in the case of Assam Sora data recorded from the Lamabari region. In the case of Assam Sora data recorded from the Lamabari region, it is observed that average *f*0 differences in the first and second syllable do not show a similar pattern. Figure 8 shows that the average *f*0 of the first and second syllable of Lamabari Assam Sora speakers is almost the same. This indicates that Assam Sora speakers of the Lamabari region are not using *f*0 differences for differentiating word stress in disyllabic words. The factors affecting this variation are yet to be determined and therefore, we argue that Assam Sora speakers of the Lamabari region are relying more on the other two cues of word stress differentiation namely vowel duration and vowel intensity.

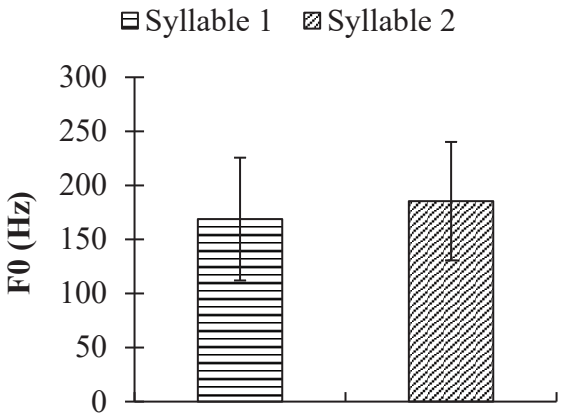

Figure 8-7: Average F0 Assam Sora (Koilamari)

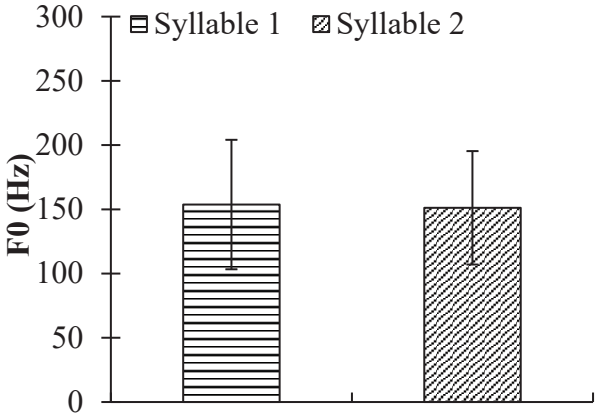

Figure 8-8: Average F0 Assam Sora (Lamabari)

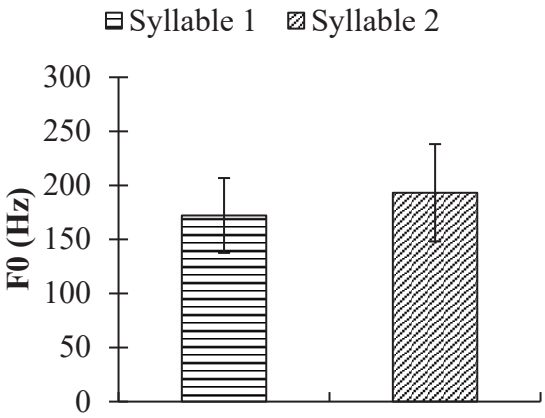

Figure 8-9: Average F0 Orissa Sora (Raiguda)

Lastly, the analysis of average vowel intensity reveals that average vowel intensity is consistently higher in the second syllable in Orissa Sora as well in Assam Sora. Figures 10, 11, and 12 show the average vowel intensity differences between the first and second syllables in Orissa Sora and Assam Sora.

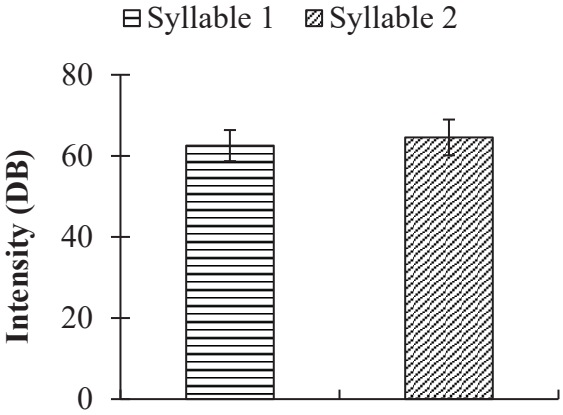

Figure 8-10: Average vowel intensity Assam Sora (Koilamari)

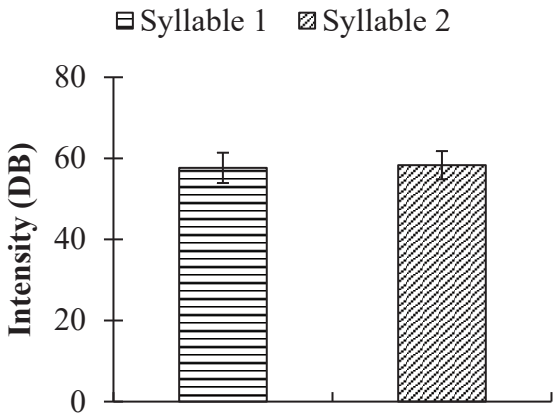

Figure 8-11: Average Vowel Intensity Assam Sora (Lamabari)

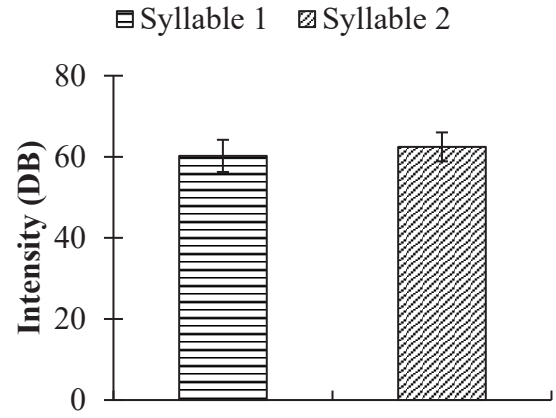

Figure 8-12: Average Vowel Intensity Orissa Sora (Raiguda)

Thus, from the analysis of word stress in Orissa Sora and Assam Sora, it is revealed that the second syllable is always longer and louder than the first syllable. Similarly, it is revealed that the second syllable is generally pitched higher than the first syllable except in the case of Assam Sora speakers living in the Lamabari region. Hence, this study proposes that

phonetic realisation of word stress in Orissa Sora is similar to Assam Sora with a minor exception that requires further examination.

4.3 Stops

Ramamurti (1938), Stampe (1965), and Anderson and Harrison (2008a) suggest that Orissa Sora has six oral stop phonemes /p, b, t, d, k, g/ and that /t/ and /d/ are asymmetric in Orissa Sora such that while /t/ is a voiceless dental stop, /d/ is a voiced alveolar stop in Orissa Sora. During the field study, it has been observed that Assam Sora also has the same six stop consonants. However, the data in this study reveals that while /t/ is consistently produced as a voiceless dental stop, /d/ is produced both as voiced alveolar stop and voiced dental stop. It appears that two voiced stops [d] and [d̪] occur in free variation in Assam Sora. Significantly, this variation is also observed in the Orissa Sora speech data recorded in this study. This indicates that the voicing asymmetry between /t/ and /d/ is often neutralized in Assam Sora as well as in Orissa Sora whereby, the voiced counterpart /d/ generally becomes similar to the voiceless dental stop consonant /t/. Subsequently, observation of the VOT measurements suggests two variations of the six oral stops in Orissa Sora and Assam Sora. Figures 13-15 represent the VOT measurements of the six-stop consonants in Orissa Sora and Assam Sora.

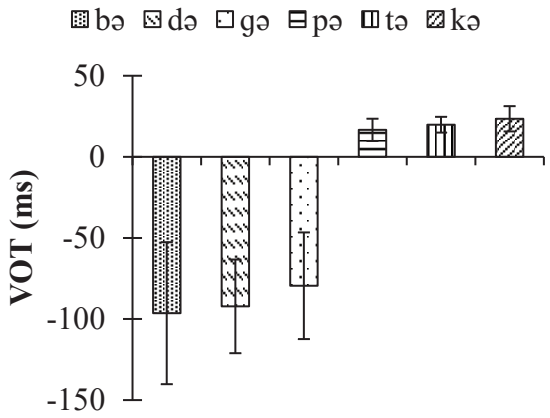

Figure 8-13: VOT Assam Sora Stops (Koilamari)

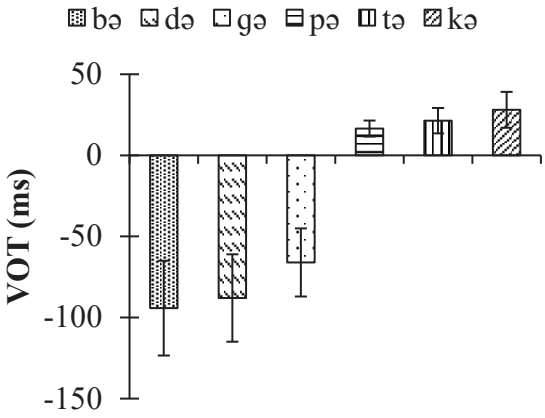

Figure 8-14: VOT of Assam Sora Stops (Lamabari)

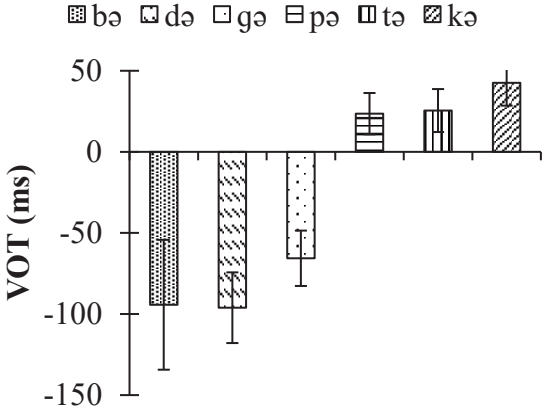

Figure 8-15: VOT of Orissa Sora Stops (Raiguda)

From Figures 13-15 it is observed that in the voiceless stop series, the velar stop /k/ always has the longest VOT, and the labial stop /p/ always has the shortest VOT, and this pattern is consistent in Orissa Sora as well as in Assam Sora. This indicates that VOT measurements in Orissa Sora and Assam Sora can also provide the cue for identifying the place of

articulation feature of the voiceless oral stops. Also, the place of articulation feature is likely to be similar in both Assam Sora and Orissa Sora. Secondly, the VOT of voiced stops in Assam Sora shows that the voiced velar stop /g/ always has the longest VOT and the voiced labial stop /b/ always has the shortest VOT. Whereas, VOT of voiced stops in Orissa Sora shows that while voiced velar stop /g/ still has the longest VOT, the shortest VOT is found in the voiced dental stop /d/ instead of the voiced labial stop /b/. This difference in the VOT values of the voiced stops in Assam Sora and Orissa Sora may be due to the free variation between voiced dental and voiced alveolar stop in both Orissa Sora and Assam Sora, but a verification of that will require further analysis.

5. Discussion

Analysis of vowel phonemes, word stress, and stop consonants in Orissa Sora and Assam Sora in this study has suggested that phonetic features of vowels, word stress, and voiceless stop consonants of Assam Sora are similar to Orissa Sora. This indicates that synchronically Assam Sora and Orissa Sora are very similar for the three phonetic properties examined in this study. These findings are significant, considering the arguments found in the literature on transplanted languages. While transplanted languages are argued to be affected by language contact, Assam Sora data reveals that transplanted Assam Sora has managed to preserve its phonetic peculiarities. This study provides evidence that phonetic features of the transplanted Assam Sora are preserved even after being separated from its place of origin. The field studies reveal that, after their migration, Assam Sora speakers have stayed inside the tea garden territory only. Also, the tea garden management has been very vigilant in preventing contact between the indentured labourers and other language communities in the vicinity. Hence, under such circumstances, Assam Sora speakers remained an isolated community in Assam even after several years of their migration to Assam. This has forbidden the chances of language contact between Assam Sora speakers and the speakers of other languages in Assam, and therefore, it was possible for Assam Sora speakers to preserve their unwritten language even after a hundred years of migration to an entirely different geographical location.

References

Anderson, Gregory D.S., Toshiki Osada, and K. David Harrison. "Ho and the other Kherwarian Languages". In *Munda Languages,* edited by Gregory D.S. Anderson, 195-255, London & Newyork: Routledge, 2008.

—, and K.David Harrison. "Sora." In *Munda Languages,* edited by Gregory D.S. Anderson, 299-380, London & Newyork: Routledge, 2008.

—, and K.David, Harrison. "Remo (Bonda)." In *Munda Languages,* edited by Gregory D.S. Anderson, 557-632, London & New York: Routledge, 2008.

—, and K. David Harrison. "Sora talking dictionary". *Living Tongues Institute for Endangered Languages*. Retrieved 2017-01-06, from http://sora.talkingdictionary.org

Hayes, B. "Iambic and trochaic rhythm in stress rules." *Annual Meeting of the Berkeley Linguistics Society* 11(1985): 429-446.

Donegan, P. J., *The phonology of Sora vowels*. Paper presented at Conference on Munda Languages and Cultures, University of Chicago, Chicago, 1981.

Donegan, P. J. "Rhythm and vocalic drift in Munda and Mon-Khmer". *Linguistics of the Tibeto-Burman Area*, *16*, no.1 (1993):1–43.

Donegan, P. J., and D.Stampe. "South-East Asian Features in the Munda languages: Evidence for the analytic-to-synthetic drift of Munda". *Annual meeting of the Berkeley Linguistics Society* 28, (2002):111–120.

Donegan, P. J., and David Stampe. "Rhythm and synthetic drift of Munda". In *Yearbook of SouthAsian Languages and Linguistics,* edited by Rajendra Singh, 3-36, Berlin, Mouton, 2004.

Ghosh, A. "Santali." In *Munda Languages,* edited by Gregory D.S. Anderson,11-98, London & New York: Routledge, 2008.

Holliday, J. J., & Kong, E. J. "Dialectal variation in the acoustic correlates of Korean stops." *Proceedings of the 17th International Congress of Phonetic Sciences*, (2011): 878-881.

Horo, Luke and Priyankoo Sarmah. "Acoustic analysis of vowels in Assam Sora." In *Northeast Indian Linguistics 7,* edited by L. Konnerth et al., 69-86. Canberra: ANU, 2015.

Jenny, M., Weber, T., and Weymuth, R. "The Austroasiatic Languages: A Typological Overview." In *The Handbook of Austroasiatic Languages*. *Volume I*, edited by Mathias Jenny and Paul Sidwell, 13-143. Amsterdam: Brill, 2015.

Kerswill, P. “Migration and language”. In *Sociolinguistics/Soziolinguistik. An international handbook of the science of language and society*, edited by Klaus Mattheier, Ulrich Ammon, and Peter Trudgill, 1–27. Berlin, De Gruyter, 2006.

Mohanty, P. “Loss of /o/ in Kui, Sora, and Oriya: A clue for a sub-linguistic area. In *Languages of Tribal and Indigenous Peoples of India: The Ethnic Space*, edited by Anvita Abbi, 109-119, Delhi: Motilal Banarasidas, 1997.

Ramamurti, G. V. *A manual of the Sora (or Savara) language*. Madras: Government Press,1931.

Ramamurti, G. V. *Sora-English dictionary*. Delhi: Mittal Publication, 1938.

Stampe, D. “Recent work in Munda Linguistics”. *International Journal of American Linguistics*, 31, no.4, (1965):332–341.

Trudgill, P. “Language contact and dialect contact in linguistic change.” *Dialektkontakt, sprakkontaktochsprakforandringi Norden: foredragfranettforskarsymposium*, (1994):13–22.

Williams, D., & Escudero, P. “A cross-dialectal acoustic comparison of vowels in Northern and Southern British English”. *The Journal of the acoustical society of America*, *136*, no.5, (2014): 2751-2761.

Zide, A. R. “*A reconstruction of Proto-Sora-Juray-Gorum phonology.”* Ph. D. dissertation, University of Chicago, 1982.

CHAPTER NINE

NOUN MORPHOLOGY IN KORKU

SHAILENDRA MOHAN

1.0 Introduction

The aim of this paper is to describe the nominal morphology of the Korku language spoken in India. Korku is a North Munda language, the westernmost language of the Austro-Asiatic phylum. It is spoken mainly in a vast area of Maharashtra and Madhya Pradesh states in India. Korku language speakers are mainly found in Nimar, Betul, Khandwa, Hoshnagabad, and Chhindwara districts of Madhya Pradesh and in Amaravati, Akola, and Buldana districts of north Maharashtra. According to the 2001 census reports, the total number of Korku speakers is 5,74,481. Through migration, Korku speakers are also found in Mumbai, Pune, Indore, Bhopal, and other cities in India.

The name of the tribe and the language was known as *Kurku*, which is now commonly referred to as *Korku.* The root *koro ~kuru* 'man or member of the Korku community takes the animate plural marker '*-ku*' to mean 'person/ member of Korku community'. The Korku language speakers are almost all bi-/multilingual, especially men, speaking the regional variety of Hindi in Madhya Pradesh and Marathi in Maharashtra. The percentage of bilingualism according to the 1991 census is 58.94. Korku has only one variety named Muwasi (Mowasi) which is spoken in the Chhindwara district of Madhya Pradesh.

The Census of India, 2001 reports the total number of Korku language speakers as 5,74,481 which includes proper Korku language speakers as 5,41,880, Muwasi as 29,288, and others as 3,313. The category "others" is mainly found in Maharashtra and Madhya Pradesh.

2.0 Noun Morphology in Korku

Noun forms in Korku show the distinction of number and case. Nouns in Korku are either root stems or derived stems i.e. derived from various word roots by morphological processes.

2.1 Number

Korku language shows the three-way distinction of numbers i.e. singular, dual, and plural. Singular nouns are unmarked, dual and plural are marked by *{-kiɲ}* and *{-ku}* in Korku respectively. Inanimate nouns in Korku don't inflect for number; the category of number is usually expressed with the help of quantifiers like two, three with inanimate nouns in Korku. Thus, animate nouns referring to both humans and non-human animate take both dual and plural markers in Korku.

Animate human nouns:

	Singular	Dual	Plural
(1)	*/poyra/* 'boy'	*/poyra- kiɲ /* 'two-boys'	*/poyra- ku/* 'boys'
	/tərəi / 'girl'	*/ tərəi - kiɲ /* 'two-girls'	*/ tərəi - ku /* 'girls'

Animate non-human nouns:

	Singular	Dual	Plural
(2)	*/kaku/* 'fish'	*/kaku- kiɲ /* 'two-fishes'	*/kaku- ku/* 'fishes'
	/minu/ ‘cat’	*/minu-kin/* ‘two cats’	*/minu-ku/* ‘ cats’

Inanimate Nouns:

In Korku only animate nouns inflect for dual and plural while countable inanimate nouns take numerals or quantifiers to express plurality. Non-countable inanimate nouns take quantifiers to express plurality.

Inanimate countable noun:

(3) /*ura*/ 'house' / *bari ura*/ ' two houses' /*ghoec ura*/ 'many houses'
/* ura- kiɲ/ */ura-ku/

(4) / *ɖa*/ 'water' /*gʰonec- ɖa* /'lots of water''
/*kuĩ*/ 'well' /*bʰ əyan kuĩ*/ 'many wells'
/ *bʰəyan kuĩ-n ɖa*/ 'the water from/of many wells'

2.2 Gender

Gender in Korku is not a part of the grammatical system itself. Gender distinctions in the Korku language are expressed by the addition of words meaning 'male' and 'female'.

(5) /*sitta*/ ' male dog' /*nali sitta*/ or /*japai sitta*/ ' female dog'
/*mara*/' male peacock' / *japai mara*/ ' female peacock'

The unmarked form is masculine in Korku and the female form is derived by adding *{japai}*.

Some lexical items, especially some kinship terms in Korku, show gender distinction /-i/ as feminine and /-a/ used as masculine markers. */-i/* and */-a/* gender distinction in lexical items is borrowed from Indo-Aryan languages spoken in the area.

(6)	masculine	gloss	feminine	gloss
	/*poyra*/	' boy'	/*tərəi*/	'girl'
	/*tiya*/	'wife's younger brother'	/*ɖukri*/	'old female'
	/*nawra*/	'bride groom'	/*neuri*/	'bride'

In Korku, */-je*/is used to mark the feminine gender, and /- *ʈe*/ is marked for masculine gender, especially in kinship terms. It was also reported in Nagaraja (1999).

(7)	*/kon/*	/'son'	*/kon-je*/	'daughter'	/*kon- ʈe*/	'male son'
	/*boko*/	'younger brother'	/*boko-je*/	'younger sister'	/*boko-ʈe*/	'younger brother'

3.0 Case marking in Korku

The cases and corresponding case markers found so far in Korku are discussed as follows:

Nominative case:

It marks the grammatical subject of a sentence. It is not overtly marked in Korku. It expresses subjects in a bare stem form without any overt case marking after it. Unlike other languages in India, there is no agreement between subject and verb in the Korku language. The nominative case marks both agents of transitives and subjects of intransitives.

(8) *poriya* *siɲ=ke* *mama-lakken*
Boy. SG tree- ACC cut- PROG
'The boy is cutting the tree'

(9) *iɲ* *giṭij- lakken*
I sleep- PROG
'I am sleeping'

Objective case:

The objective case is signalled by the accusative suffix *–ke/kʰe* in Korku. The accusative marker optionally appears on the object and generally marks themes and patients. The appearance of the accusative has to do with the factor 'animacy' i.e. only animate objects take the objective case marker.

(10) *iɲ* *ambe* *jojom-ba*
I mango eat-FIN/FUT
'I eat mango'

(11) *ape* *sitta=ke* *munḍa-ke-nec*
You.PLU dog-ACC hit-PST-OBJ MAR
'You hit the dog'

In the above examples, one can see that in example (10) vs (11), the noun subcategorises for objective case on the noun complement if it is an animate noun.

Dative case:

The dative case marker can be used to mark the recipient only. In Korku, the dative form is used to mark the recipient and the patient/ theme.

(12) *ɨɲ* *ɖiku=ke* *kitab* *ji-ke*
I they- DAT book give-PST
'I gave them the book'

(13) *ape* *i ɲ=kʰe* *sitta* *ji-ke-nec*
You.PLU I-ACC dog give-PST- OBJMAR
'You gave me a dog'

(14) *ɨɲ=en* *kʰusi* *ɖan*
I-DAT happy PST
'I was happy'

In Korku *-en/-n* marks the dative subject/ experiencer subject. It appears that Korku does not have the dative marker. The */=ke or =kʰe/* marker is a true primary object marker.

Benefactive case: In Korku, benefactive is marked by */ =gʰelya/*.

(15) *ɖic* *ram= gʰelya* *kitab* *sa-le*
he ram-BEN book buy-PERF
'He has bought a book for Ram'

Genitive case**:**

The Genitive case is the case that observes the relationship of something with a noun or a pronoun. In Korku, it is marked by *-a/ya/ga.*

(16) *ɨɲ=ya* *konje*
I-GEN daughter
' My daughter'

(17) *porya=ga* *jumu*
boy-GEN name
'Boy's name'

(18) *am=a* *konje*
you-GEN daughter
'Your daughter'

Instrumental Case:

The instrument marker marks the instrument by means of which an action is performed. Korku has an instrument case marker that has instrument and manner functions.

(19) *ḍiku* *caku-=ʈen* *jilu* *ma-wen*
they knife-INST meat cut-PERF
'They cut the meat with the knife'

(20) *ḍic* *jor= ʈen* *sarup-en*
he fastness-INST run-PERF
'He ran fast'

Comitative case:

Korku marks the comitative case by clitic '=*gon*'.

(21) *ḍic* *iɲ=ya=gon* *haʈi* *ol-en*
he I-GEN-COM market go-PERF
'He went to the market with me'

The commutative is marked with a fused postpositional structure and has the shape *N-GEN-COM*.

4.0 Spatial Markers

Four important spatial markers are widely discussed generally. The case labels for these cases are ablative (source), perlative (path), allative (goal), and locative (static location).

Ablative, Perlative, and Allative cases:

Korku marks the source of motion and the path by the same form. The ablative and perlative cases are marked by a similar form.

(22) *iɲ jamud= ʈen hec-ken*
I Jamud- ABL come-PERF
'I came from Jamud'

(23) *iɲ ura= ʈen hec-ken*
I house-through come-PERF
'I came through the house'

Korku also marks the static and dynamic ablatives in similar ways. The end point marker, i.e. allative, is marked by *=ʈay* in Korku.

(24) *ɖic jəlgaon= ʈen jamud= ʈay sərəko haru-ke*
he Jalgaon-ABL Jamud- LOC road make-PERF
'He built a road from Jalgaon to Jamud'

(25) *ɖic jəlgaon- ʈen jamud- ʈay sarup-ke*
he Jalgaon-ABL Jamud- LOC run-PERF
'He ran from Jalgaon to Jamud'

Locative case**:**

Korku marks the locative -on with */=liyen/* and locative -in with */=ʈalan/.*

(26) *pustək ʈebəl=liyen ʈaka*
book table-LOC be
'The book is on the table'

(27) *ram ura=ʈalan=tic ke-nec*
book house-in-LOC be- OBJMAR
'Ram is in the house'

5.0 Derivations of Nouns

Agentive nouns are derived by adding /-mit^{h}ac/ and /-minij/ suffix to both verbal root and noun root in Korku,

(28) */ ura/* 'house' /*ura-mit^{h}ac*/ or / *ura-minij*/ 'house owner'
/*heje/* 'to çome' /*heje- mit^{h}ac*/ or / *heje-minij*/ 'comer'
/*ol*/ 'to write' /*ol -mit^{h}ac*/ or / *ol-minij*/ 'writer'

The above observation leads to the debate about the distinction between noun vs verb root in Munda languages. It has been cited that Mundari, a Munda language, is an example of a language without word classes, where a single word can function as a noun, verb, adjective, etc. according to the context (Hoffman, 1903). Bhat (1997:249) concludes for Munda that "the noun-verb distinction can only be viewed as a functional one". Later, Evans and Osada (2005) argue that Munda languages clearly distinguish nouns from verb roots. (For detailed discussion refer to *Linguistic Typology* 9, 2005).

6.0 Plural Agreement with Adjective

Korku exhibits plural agreements with the adjectives. In Korku, plural agreement is shown with infixation. In Korku; */-pe-/* is the infix to show the plural agreement with the noun.

(29)	/kʰa-pe-t ɖʰega/	'big stones'
	/ kʰa-pe-t kon-ku/	'big boys'
	/kʰat ɖʰega/	'big stone'
	/ kʰat kon/	'big boy'

7.0 Numerals in Korku

This section deals with the cardinal numeral expressions in the Korku language.

Numerals from 2-10

(30)

Numerals	Korku	**Numerals**	**Korku**	**Numerals**	
2	*bari*	12	*gel ɖo bari (10 and 2)*	30	*isa ɖo gel (20 and10)*
3	*apʰei*	13	*gel ɖo apʰei (10 and 3)*	31	*isa gel ɖo mya? (20+10 and1)*
4	*upʰun*	14	*gel ɖo upʰun (10 and 4)*	40	*bari isa (2x20)*
5	*monoy*	15	*gel ɖo monoy (10 and 5)*	50	*bari isa ɖo gel (2x20 and 10)*
6	*turui*	16	*gel ɖo turui (10 and 6)*	51	*bari isa ɖo gel mya? (2x20 and 10 +1)*

7	*yei*	17	*gel ḍo yei (10 and 7)*	70	*ap^{h}ei isa ḍo gel (3x20+10)*
8	*ila*	18	*gel ḍo ila (10 and 8)*	72	*ap^{h}ei isa ḍo gel bari (3x20+10+2)*
9	*a:rai*	19	*gel ḍo a:rai (10 and 9)*	100	*mya? seddi 1x100)*
10	*gel*	20	*isa (20)*	112	*mya? seddi gel ḍo bari 1x100+10 and 2)*

Korku cardinal numerals are relatively uniform among the speakers. There is a slight regional variation in forming cardinal numbers in Korku of the Amravati district (see Nagaraja 1999:82), but the system remains the same. An examination of the set of numerals from '11 to 19' reveals that the numerals are formed by the addition of '10' and a unit numeral in Korku. There seem to be no monomorphemic forms above 10 in Korku. Korku exhibits bi-morphemic patterns to form '10' and above numerals with a base of '10'. They form their complex numerals by the pattern 'base unit 10+ R'. The formation is consistent with Dravidian language patterns where base unit 10 + remainders (R) is added. Dravidian languages form their '11 to 19' cardinal numbers as '10+3=13', for example, Kannada forms 'thirteen' as *hadimuru* i.e. 10+3 (Andronov, 1976), while Indo-Aryan languages form cardinal numbers from '11 to 19' in an opposite pattern i.e. 'thirteen' is formed as '3+10'. Even English, for instance, combines the elements *six* '6' and *ten* '10' to form *six-teen* '16' employing the addition of the two numerical values. Norman Zide (1978:1) observes that "Proto-Munda had both a duodecimal and a vigesimal system".

An examination of the set of complex numerals from '30 to 100' reveals the pattern of the vigesimal system, and they are predominately formed by multipliers (1-5) x base unit '20'+ remainders again. Korku is consistent in forming the complex numerals with a base unit of '20' i.e. it is a vigesimal system.

8.0 Interrogative Words

Interrogative words are words like the English *who, what, where, when,* etc., as they are used at the beginning of questions. They include a set of interrogative pronouns, interrogative adverbs, and indefinite forms. The forms in Korku are presented below:

(31)	Interrogative words	Korku
	'who'	*je*
	'whose'	*je-ga*
	'with whom'	*je-gon*
	'where'(exact location)	*tungən*
	'where'(inexact location)	*ɖingən*
	'how many'	*coʈo*
	'when'	*cola*
	'how'	*cupʰar*
	'why'	*cuja*
	'what'	*cuc*

9. Postpositions in Korku

Korku postpositions are listed below:

List of postpositions	Korku
'in front of house'	*ura səmma*
'inside a river	*gaɖa ʈalan*
'far from house'	*ura- ʈen*
'behind the house'	*ura gaʈəu*
'near the house'	*ura meran*
'above the house	*ura-liyen*
'in direction to house'	*ura konen*
'next to house'	*ura-ga baju*
'below the table'	*tebəl- iʈan*

10. Conclusion

In this study, the nominal morphology of the Korku language has been described. It is hoped that the description will help to fill the gap that exists about the knowledge of this language, as well as to provide data for the comparative study of the other South Asian languages.

Abbreviations

ABL	Ablative Case marker
ACC	Accusative Case marker
BEN	Benefactive
COM	Commitative
DAT	Dative Case marker

FIN/FUT	Finite/ Future marker
GEN	Genitive Case marker
INST	Instrumental
LOC	Locative Case marker
OBJMAR	Object marker
PERF	Perfect Aspect marker
PLU	Plural
PST	Past marker
PROG	Progressive Aspect marker

Colophon

I am very much indebted to Prof. Masato Kobayashi who helped me with his valuable suggestions for the improvement of this paper. I am thankful to my Korku informants, Tibu Kazde, Silal sir, and Jawarkar sir. I am thankful to the ICSSR-JSPS research grant to conduct the fieldwork.

References

Anderson, Gregory D.S. *The Munda Verb. Typological perspectives.* Berlin: Mouton, 2006.

—, ed. *The Munda Languages.* London / New York: Routledge, 2008.

Andronov, M.S. "Dravidian Numerals: An Etymological study." *International Journal of Dravidian Linguistics* 5, no.1 (1976):5-15.

Bhatt, D.N.S. "Noun-Verb distinction in Munda languages." In *Languages of Tribal and Indigenous Peoples of India: The Ethnic Space*, edited by Anvita Abbi, 227-251. Delhi: Motilal Banarsidas, 1997.

Census of India. Table C-16, Office of the Registrar General, India, 2001.

Census of India. Table C-16, Office of the Registrar General, India, 1991.

Evans, N & T. Osada. "Mundari: The myth of language without word classes." *Linguistic Typology* 9 (2005), 351-390.

Hoffman, John. *Mundari Grammar*. Calcutta: The Secretariat Press, 2003.

Mohan, Shailendra. "Noun morphology in Nihali and Korku, a comparative account." *Bulletin of the Deccan College Research Institute* Vol. 75 (2015): 185-194.

Nagaraja, K.S. *Korku Language. Grammar Texts, and Vocabulary.* Tokyo: Institute for the Study of Language and Cultures of Asia and Africa, 1999.

—. "Nahali and Korku: A comparative note". *Bulletin of the Deccan College Research Institute*, Vol. 66/67(2006-07): 315-322.

Zide, Norman. *Studies in the Munda Numerals*. Mysore: Central Institute of Indian languages, 1978.

CONTRIBUTORS

Gregory D.S. Anderson is a founder and director of Living Tongues Institute for Endangered Languages. He has published widely in the fields of descriptive grammar, historical linguistics, morphology, and verb typology, and endangered and minority languages of Siberia, India (Munda & Tibeto-Burman), Papua New Guinea, Native North America, and Africa.

Nathan Badenoch is the Director of the Kyoto University North American Center. Through his project "Beyond Iconicity: Typological Research in Expressives", a Grant-in-Aid project from the Japanese Society for the Promotion of Science, he has worked on expressives in South, East, and Southeast Asia. He is co-author of *A Course in Mundari* (2015) with Toshiki Osada, Madhu Purti, and NishaantChoksi, as well as the forthcoming *Dictionary of Mundari Expressives*, co-authored with Toshiki Osada and Madhu Purti.

Nishant Choksi is an assistant professor in the Department of Humanities and Social Sciences at the Indian Institute of Technology, Gandhinagar. He has worked extensively on Munda languages and scripts such as Santali and Mundari. He has published articles in peer-reviewed journals such as the *Journal of Linguistic Anthropology, Modern Asian Studies,* and *International Journal of the Sociology of Language*. He also co-authored *A Course in Mundari* (2015) with Nathan Badenoch, Toshiki Osada, and Madhu Purti, with whom he is now collaborating on a project on Mundari Expressives.

Martin Everaert is an emeritus professor of Linguistics at Utrecht University. He works primarily on the syntax-semantics (anaphora) and the lexicon-syntax (idioms, argument structure) interface. His other areas of interest are typology and language evolution. He is, a.o., co-editor of the *Wiley-Blackwell Companion to Syntax* and member of the editorial boards of *Linguistic Inquiry* and *Journal of Linguistics*.

Arun Kumar Ghosh (pen name Arun Ghosh) is a Retired Professor of Bengali, The University of Burdwan, West Bengal. and former ICCR Tagore Chair, Centre for Modern Indian Studies, Georg-August University, Goettingen, Germany (2013-15). He was a Research Fellow/ Scientist, Department of Linguistics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany in 2005, 2006, 2007, and 2010. Engaged in research on the Munda languages, especially Santali, Gutob, Remo, and Gta' since 1980. He has penned several books and articles on the Munda languages published by renowned publishers and in national and international journals. His most recent publication is "Case, Agreement and Argument Marking in Munda" *IJDL*, 48:1, 2019.

Luke Horo is an institute fellow and postdoctoral researcher for the South Asia region with Living Tongues Institute for Endangered Languages. He specialises in phonetics of Munda languages in India and recently completed his Ph.D. dissertation "A Phonetic Description of Assam Sora", from the Department of Humanities and Social Sciences, Indian Institute of Technology Guwahati. His publications include acoustic analysis of vowels, consonants, and lexical stress of the Sora language.

Bikram Jora is a Regional Coordinator of South Asia with the Living Tongues Institute for Endangered Languages. His research mainly focuses on morpho-syntax, negation, descriptive grammar, and language acquisition of Munda languages (Kherwarian languages). He is an expert field linguist and has documented many Munda and Kho-Bwa languages spoken in India.

Masato Kobayashi is a Professor of linguistics at the University of Tokyo. He is a linguist working on Indian languages of the Indo-Aryan, Dravidian, and Munda groups, and is the author of “Historical Phonology of Old Indo-Aryan Consonants (Tokyo 2004), Texts and Grammar of Malto (Vizianagaram 2012), and The Kurux Language with Bablu Tirkey, The Hague 2017)”.

Anish Koshy is an Assistant Professor (Linguistics) in the Department of Linguistics and Phonetics at the English and Foreign Languages University, Hyderabad. He works on lesser-studied languages of India, both from a descriptive as well as a comparative typological perspective. His doctoral thesis is on the Austroasiatic languages of India.

Shailendra Mohan is Professor in Austroasiatic Linguistics at the Deccan College PG& RI (Deemed University), Pune. His areas of specialization include language typology, language contact, and endangered languages. His primary research interests are Munda linguistics, and the language isolates Nihali**.**

Toshiki Osada is an Emeritus Professor at the Research Institute for Humanity and Nature (RIHN), Kyoto, Japan, and the leader of the Indus Project since 2012. He is a linguist working on South Asian languages, especially Munda languages spoken in Jharkhand. He has published *A Reference Grammar of Mundari* (1992) and several papers on Munda languages. He also worked on Indus Civilization and edited *Current Studies on Indus Civilization* Vols. 1-9.

Madhu Purti was born in the remote Mundari village, Kewra, in the dense forest in the state of Jharkhand, India in 1963. She graduated from Gossner College in Ranchi and is an expert in traditional Munda culture and Mundari folksongs. She worked with Toshiki Osada, Nathan Badenoch, and Nishaant Choksi as a native speaker of Mundari. She is a co-author of *Dictionary of Mundari Expressives* (forthcoming).

Priyankoo Sarmah is an Associate Professor of Linguistics at the Indian Institute of Technology Guwahati. His specialization is in the phonetics and phonology of the languages of North East India. He is particularly interested in the tone languages of the region and in language technology development in under-resourced languages. His publications primarily deal with the vowels and tones in the languages of North East India.

Karumuri V. Subbarao was Professor at Delhi University (1974-2005) and also, Radhakrishnan Chair Professor in Humanities at the University of Hyderabad (2011-12). He was a Visiting Professor at the University of Hamburg, Germany, and Tokyo University of Foreign Studies, Japan. His main areas of research include the study of the syntactic typology of South Asian languages, Munda languages, Tibeto-Burman languages, and Dravidian languages. He is the author of South Asian Languages: A Syntactic Typology. Cambridge. New York and; Delhi: Cambridge University Press. He is an Elected Member for the life of the Linguistic Society of America in 2003 and the Linguistic Society of Nepal in 2005.

INDEX